

# **Auditory Inspection Time and Intelligence**

---

**Ian J. Deary**

**Doctor of Philosophy**

**University of Edinburgh**

**1992**



To my family



## **Declaration**

**This thesis is my own composition, and the work presented in it is my own.**

## **Acknowledgements**

I thank Peter Caryl for providing the computer programming expertise required to implement the inspection time tests, for discussions on the studies and for helpful comments on earlier drafts of this thesis. I thank Chris Brand for his generous provision of time for discussion on inspection time and for his helpful comments on drafts of the enclosed chapters. I am grateful to Vincent Egan for his help in running the experiments reported in Chapters 4 and 5. My thanks go to Brian Head and his fellow teachers for helping to provide subjects for the longitudinal study. The design and construction of the auditory inspection time apparatus was the work of the late David Wight and his contribution to this thesis is warmly appreciated.

## **Abstract**

This thesis studied the association between auditory inspection time (AIT) and psychometric measures of verbal and non-verbal cognitive abilities. I review attempts to search for basic information processing components that predict intelligence (Chapter 1), attempts to relate auditory processing speed to intelligence (Chapter 2), and attempts to relate acuity of sensory discrimination to intelligence (Chapter 3). These reviews establish certain essential requirements for a plan of research on auditory inspection time.

Chapter 4 described the development of a modified AIT test. In a study of 120 undergraduates, the modified AIT test showed improved subject performance characteristics over previous AIT tasks, and AIT thresholds had low to moderate correlations with visual IT thresholds and with verbal and non-verbal cognitive ability scores. Chapter 5 described two studies. Study 1 included 84 undergraduates and showed that the AIT test had a very high split-half reliability and that about two-thirds of subjects who could perform the AIT task had response performance curves which fitted a cumulative normal ogive. The association between AIT and verbal ability appeared stronger than the AIT-non-verbal ability association in 34 of the subjects; this was also found in Study 2 which tested 119 11-year-olds. Unspeeded pitch discrimination showed a small but significant association with verbal ability in children but not in undergraduates. Results from neither study supported the suggestion that pitch discrimination was the basis for the AIT-cognitive ability association.

Chapters 6 and 7 examined the associations among AIT, unspeeded pitch discrimination and an auditory backward masking recognition task which was dubbed the 'Raz' task. It was found that all three tasks were reliable, prone to practice effects and showed high intercorrelations. The AIT and Raz tasks appeared to share common variance not related to pitch discrimination. In a confirmatory factor analysis of over 100 13-year-olds latent

variables from the three auditory tests representing auditory processing speed and pitch discrimination both had significant associations with a factor common to verbal and non-verbal intelligence, though speed was the more important factor.

Chapter 8 reported the results of a longitudinal study of AIT and cognitive ability in over 100 children from age 11 to age 13. Using structural modelling techniques to create competing causal models and then testing these for goodness-of-fit to the data, some support was found for the suggestion that auditory processing abilities at age 11 might have a causal influence on later verbal and non-verbal abilities rather than the converse.

Chapter 9 provided a thematic resume of the studies conducted in the thesis. It was concluded that the corrected AIT-cognitive ability association was in the region of -0.5, and that some progress had been made in explaining this association. In addition, a strong plea was made for AIT and visual IT to be integrated with other models of auditory and visual information processing which exist. Suggestions were made for future research on auditory and visual processing and intelligence.

## **Contents**

### **Chapter One: Intelligence: a descriptive or explanatory concept?**

1.1	Tautology in reasoning about human intelligence	1
1.2	Resolving psychometric models of cognitive ability	5
1.3	Causes and consequences of cognitive ability test scores	6
1.4	The 'biology' of intelligence	9
1.5	Psychological 'components' underlying psychometric test performance	14
1.6	Inspection time	24

### **Chapter Two: Intelligence and auditory information processing**

2.1	Auditory inspection time and intelligence	35
2.2	Other studies of auditory information processing and intelligence	44
2.3	Conclusions and a research plan	50

### **Chapter Three: Sensory discrimination and intelligence: historical review and re-analysis**

3.1	Introduction	54
3.2	Current conceptions of the early studies	55
3.3	The studies of Spearman and Burt	65
3.4	Intelligence and sensory discrimination after Spearman and Burt	75
3.5	Conclusions	85

### **Chapter Four: An improved auditory inspection time task: associations with cognitive ability and visual inspection time**

4.1	Introduction	88
4.2	Methods	95
4.3	Results	101
4.4	Discussion	107

### **Chapter Five: Pitch discrimination, auditory inspection time and cognitive abilities**

5.1	Introduction	114
<b>Study 1</b>		
5.2	Method	116
5.3	Results	118
5.4	Discussion	128
<b>Study 2</b>		
5.5	Subjects	132
5.6	Method	132
5.7	Results	135
5.8	Discussion	141
5.9	General discussion of studies 1 and 2	144

## **Chapter Six: Auditory inspection time, cognitive ability and the 'Raz' task- I**

<b>6.1</b>	<b>Introduction</b>	<b>148</b>
<b>6.2</b>	<b>Method</b>	<b>155</b>
<b>6.3</b>	<b>Results</b>	<b>159</b>
<b>6.4</b>	<b>Discussion</b>	<b>169</b>

## **Chapter Seven: Auditory inspection time, cognitive ability and the 'Raz' task - II**

<b>7.1</b>	<b>Introduction</b>	<b>174</b>
<b>7.2</b>	<b>Method</b>	<b>175</b>
<b>7.3</b>	<b>Results</b>	<b>179</b>
<b>7.4</b>	<b>Discussion</b>	<b>198</b>

## **Chapter Eight: Auditory inspection time and cognitive ability: a longitudinal study**

<b>8.1</b>	<b>Introduction</b>	<b>201</b>
<b>8.2</b>	<b>Method</b>	<b>203</b>
<b>8.3</b>	<b>Results</b>	<b>204</b>
<b>8.4</b>	<b>Discussion</b>	<b>220</b>

## **Chapter Nine: Conclusions**

<b>9.1</b>	<b>Concluding remarks concerning the studies in this thesis</b>	<b>225</b>
<b>9.2</b>	<b>Integrating inspection time research with other paradigms</b>	<b>231</b>
<b>9.3</b>	<b>Reductionistic explanations of the IT-cognitive ability association</b>	<b>241</b>
<b>9.4</b>	<b>Conclusion</b>	<b>245</b>

<b>References</b>	<b>247</b>
-------------------	------------

<b>Appendix 1</b>	<b>263</b>
-------------------	------------

<b>Appendix 2</b>	<b>266</b>
-------------------	------------

<b>Appendix 3</b>	<b>268</b>
-------------------	------------

<b>Appendix 4</b>	<b>270</b>
-------------------	------------

<b>Appendix 5: Published papers</b>	<b>271</b>
-------------------------------------	------------

# Chapter One

## Intelligence: a descriptive or explanatory concept?

*A first, extremely important, fact is that all tests of mental abilities tend to give positive inter-correlations. (Vernon, 1940, p153.)*

*The primary, or fundamental attributes of Thought, or Intelligence, have been already stated to be, Consciousness of Difference, Consciousness of Agreement, and Retentiveness. ...and the most fundamental property is the Consciousness of Difference, or DISCRIMINATION. (Bain, 1868; emphases in the original.)*

### 1.1 Tautology in reasoning about human intelligence

...if we were asking someone to explain why one factory produces more than another, we would not be satisfied with the explanation 'because it is more productive'. In this context the term 'productive' does not genuinely explain anything at all; it simply rephrases what is already known. The failure arises from the fact that while 'productivity' is a valuable *descriptive* term, it is not an effective *explanatory* concept.

...the concept of intelligence, although often utilized in ways that its users believe to be explanatory, is in fact restricted in precisely the same way, and... invoking high measured intelligence in order to explain a person's success is no more meaningful than putting forward productivity as the explanation for a factory's level of output.

(Howe, 1988a; emphases in the original.)

As illustrated by the above quotation from Howe (1988a), reasoning about the meaning of putative intelligence differences demonstrated by psychometric IQ-type tests is often held to be circular. The critics who make this charge are repeating, however unknowingly, the dictum of Edwin Boring, who was the first to remark, in 1923, that intelligence research was vulnerable to the criticism that intelligence, "is what the tests test". Until recently it has not been possible to answer this objection definitively. In this first Chapter it will be

argued that a definitive answer to the charge that intelligence is merely a descriptive term is unlikely to come from psychometric studies alone. In agreement with some of the suggestions of Howe (1988a,b), it will be argued that it is necessary to validate estimates of intelligence obtained from psychometric tests by establishing that IQ scores reliably predict success in real-life tasks and, more importantly, that IQ scores have bases in physiological processes or in basic psychological functions.

Psychometric studies of human intelligence have provided important findings which will form the bases of attempts to validate the concept of intelligence. Psychometricians have long held that, for a large random sample of the population, when such a group is tested on a variety of different mental tests, the correlation matrix that results is composed almost entirely of positive correlations. This tends to run counter to most people's intuitive conceptions of the structure of mental abilities, i.e. everyday experience tells one that some people excel at some mental tasks and do poorly on others. The positive association among most mental tests is the finding that led Spearman (1904) to propose that, due to differences in brain functioning, the normal population showed persistent and reliable differences in general intelligence (*g*). This conclusion has not proved easy to substantiate. For, while Spearman's positive manifold is a common empirical finding, its interpretation is not uncontroversial.

Confirming Spearman's discovery, recent principal components analyses of the 11 very different subtests of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) (e.g. Blaha and Wallbrown, 1982; Canavan, Dunn and McMillan, 1986; Crawford, Allan, Stephen, Parker and Besson, 1989) have demonstrated that the *g* factor extracted as the first unrotated principal component from the inter-test correlations accounts for consistently more than 50% of the between-subjects variance on the WAIS-R tests. Even the lowest subtest loadings on the *g* factor are typically around or greater than 0.6. Attempts to replace conventional IQ-style tests with non-*g* loaded mental ability tests have not proved



successful. For example, Hooper, Hooper and Colbert (1984) reported a study involving subjects of different age groups who were administered standard psychometric intelligence tests along with tests developed from Piaget's theory of formal operational thinking. The average correlation between Raven's Progressive matrices and twelve formal operational reasoning tasks was 0.53.

The ubiquity of positive correlations among tests of cognitive abilities notwithstanding, there is a real problem in the interpretation of *g* which lies in the statistical methodology that is used to demonstrate it. It is possible to choose a particular method of factor analysis and, from the characteristics of the method, fashion one's own model of intelligence. Thus, followers of Spearman's preferences such as Burt (1909-10), Jensen (1980) and Eysenck (1982) have continued to extract the general factor using methods like principal components analysis and emphasising the results of the unrotated first principal component. Others claim that this methodology tends to ensure the demonstration of a general factor which has little importance (Gould, 1981). Undheim (1981a,b) has independently arrived at a view about the structure of human mental abilities which emphasises Spearman's *g*, and Undheim has referred to his own research as leading to a "restoration of general intelligence".

The occurrence of the positive manifold of correlations among mental tests does not necessarily lead to researchers including the concept of general intelligence in their models of human intellect. Among psychometricians, the theorist most opposed to *g* was probably Guilford (1985), who tended to implement factor analytic techniques that 'ignored' *g*. Thus, he was able to construct a model of intelligence which contained 120 (lately 150) separate mental abilities though, as Eysenck (1979) pointed out, many of these were positively correlated and, if submitted to higher order factor analysis, tended to yield fewer and more general mental ability factors. The charge that Guilford's 'Structure-of-Intellect' model ignored *g* does not resolve the matter, because there are other

influential models of human abilities that do not include *g*. When oblique rotation is used in factor analysis of a sample's mental ability test scores (Kline 1991a), several relatively separate primary mental abilities emerge that are similar to those proposed by Thurstone (1938), and second order factors may be discerned, e.g. fluid and crystallised intelligence factors such as have been hypothesised by Cattell (1963).

Gustafsson (1984) explained how the implementation of different factor analytic techniques allowed researchers to arrive at such different opinions as to the importance of the general factor in intelligence,

...when Multiple Factor analysis is used with orthogonal rotation, the general factor is "rotated away," by being represented as small positive loadings in all factors. However, in interpretations of factor analytic findings, loadings of lower than 0.3 are rarely attended to, and often not even represented. It may thus be claimed that orthogonal rotations to simple structure are quite deceptive in the presence of a general factor.

If an oblique rotation is carried out, the general factor is represented as the correlations among the factors. There are two problems inherent in oblique rotations, however. One is that there are almost always small positive loadings scattered in the matrix, which cause the true correlations among factors to be underestimated. The other problem is that most oblique rotational methods allow the researcher to determine the degree of obliqueness of the solution...

The psychometric tradition has not been able to resolve the different models of intelligence (Kline 1991a, Sternberg 1990). The limitation of traditional factor analytic techniques was captured in the following quotation from Sternberg (1981),

Almost eighty years after the first presentation of Spearman's (1904) two factor theory, has anyone answered through factorial means the question of whether a general factor exists?

...[factor analysis has] failed because it has been too successful in supporting, or at least failing to disconfirm, too many alternative models of intelligence.

Referring to his belief that the concept of intelligence cannot be validated by psychometric methods, Miles (1988) stated that, "Psychometrics is at best a useable technology and has no status as a science". Although Miles' (1988) opinion of the possibilities offered by

psychometrics was more pessimistic than that of other critics of psychometric research (Howe, 1988b), the dependence of intelligence research on traditional psychometric methods is perhaps now lessening, and a resolution of different models of intelligence and escape from the charge of tautology may be possible, as this thesis will suggest.

## **1.2 Resolving psychometric models of cognitive ability**

The methods of traditional factor analysis are being replaced by latent trait methods which allow correlational data to be tested against theoretical models. Referring to the problems that traditional factor analytic methods have had in helping psychologists<sup>to</sup> choose among different models of intelligence, Gustafsson (1984) stated, "In confirmatory factor analysis, however, all these problems are avoided". Confirmatory factor analytic methods are still limited to psychologists who have acquired the necessary statistical sophistication. A non-technical introduction to the various uses of latent trait models in intelligence research was provided by Whitely (1980), and an annotated bibliography of psychological applications of structural equation modelling was compiled by Austin and Wolfe (1991). Briefly, the development of these statistical methods has led to a departure from the time-honoured procedures in intelligence research of selecting a particular method of factor analysis, feeding in a correlation matrix and allowing the results of the analysis to drive one's theoretical conceptions of intelligence. It is now possible to test hypotheses concerning the differential goodness-of-fit of models of intelligence to a set of data, i.e. conceptions of intelligence formed explicitly, and in advance of any analysis, may be tested competitively to see how well each model corresponds to the structure of the correlations within the data.

The techniques of confirmatory factor analyses have been used in this way to test the factorial structure of intelligence test batteries (Bynner and Romney, 1986; O'Grady, 1989), using statistical packages such as LISREL (Joreskog and Sorbom, 1986) and EQS (Bentler, 1989). For instance, Gustafsson (1984) administered 16 tests of cognitive ability

to 1000 subjects and tested various models of intelligence for their goodness-of-fit to the data. Interestingly, the model which was selected as having the best fit to the data was able to give support to many previous intelligence theorists. Older models of the structure of human abilities appeared as special cases in Gustafsson's (1984) hierarchical, "unifying model" of intelligence. At a fairly low level in Gustafsson's successful hierarchical model were primary factors similar to those hypothesised by Thurstone (1938) and Guilford (1985). At a higher level in the hierarchy there appeared the fluid and crystallised abilities proposed by Cattell (1963). At the peak of the hierarchy sat *g*, which was not able to be distinguished from the second order factor of fluid intelligence. With the work of Gustafsson (1984; Undheim and Gustafsson, 1987), models of intelligence which had seemed irreconcilable for so long emerged as complementary rather than contradictory. Moreover, Gustafsson's (1984) hierarchical model gave support to a similar model of human abilities proposed by Horn (1980) on the basis of a theoretical attempt to integrate different models of intelligence.

### **1.3 Causes and consequences of cognitive ability test scores**

The advances in statistical techniques notwithstanding, for as long as accounts of individual differences in intelligence are based upon the results of psychometric tests and their interrelations, many will remain unconvinced about the 'reality' or validity of cognitive ability differences (Howe, 1988b). Indeed, many have written to the effect that psychometric intelligence is just a beginning, and that the structure of test interrelations constitutes something to be explained rather than being an explanation in itself (Deary, 1988; Eysenck, 1986). Eysenck (1986), extended the Intelligence A (innate potential) and B (average level of performance) ideas of Hebb (1949), and suggested that three types of intelligence should be recognised. In Eysenck's (1986) 'new model of intelligence', IQ represented 'psychometric intelligence' that was measured by standard cognitive ability tests and was affected by education, family upbringing and cultural factors. Underlying this was 'biological intelligence', i.e. innate ability, conceptualised in terms of

biochemistry and physiology, and heavily influenced by the genes, and measurable in terms of EEG, reaction times and evoked potentials. The application of psychometric intelligence to problems encountered in the world resulted in individual differences in 'social intelligence', and its expression was seen as being affected by many factors, including health, personality, nutrition and motivation.

Sternberg (1990) called the psychometric approach to intelligence the 'geographic metaphor' and portrayed it as just one of a number of self-contained approaches that 'look inward' upon the individual and attempt to *describe* intelligence in their own terms. Other inward looking approaches included the 'computational metaphor', the 'biological metaphor' and the 'epistemological metaphor'. Sternberg's thesis was that intelligence could not be *explained* until newer, more all-encompassing models of intelligence were constructed. While each metaphor remained isolated from the others all we may expect, according to Sternberg (1990), is a number of equally valid stories concerning intelligence differences that are blind to other approaches.

Howe (1988a) argued that, if the concept of intelligence is to escape the charge of being a word which represents a thing having no substance, it must meet certain criteria which may be demanded of any putative explanatory phenomenon. In other words, intelligence must have demonstrable links to things outside of the tests which are used to estimate individual differences to be considered an explanatory concept rather than a tautologous description. Howe's (1988a) desiderata for validating the concept of intelligence were that intelligence test scores should be related to: physiological variables, basic mental processing mechanisms, the capacity to learn or remember, fundamental thinking skills, the ability to reason abstractly, the complexity of a person's cognitive functioning, mental flexibility, executive controlling functions, biological mechanisms and indications that the level of measured intelligence precisely identifies intellectual qualities possessed by an individual. Sternberg (1988) criticised this list, claiming, in contradiction to Howe (1988a), that some

of these criteria had already been established, while others were not easily operationalised. Nevertheless, a powerful editorial in the Lancet (1991) to the effect that intelligence is an illusory concept owed much to Howe's (1988a,b) arguments that the research on intelligence is an exercise in circularity, with little or no reality outside of its own terms of reference.

Despite the fact that Sternberg (1988) answered many of Howe's (1988a) criticisms of intelligence research, and despite the empirical evidence cited by Brand, Caryl, Deary, Egan and Pagliari (1991) to indicate that intelligence is not an illusory construct (Lancet, 1991), the above accounts of intelligence research have a common theme and this theme must be taken up and considered. The accounts of intelligence by Eysenck (1986), Sternberg (1990), Howe (1988a,b) and the Lancet editorial (1991) were all based on the premise that psychometric research must be validated by other considerations in order to demonstrate that intelligence is a useful or explanatory psychological concept. It is not enough to be able to demonstrate that individual differences in psychometric intelligence are reliable, as they have proved to be across several decades (Schwartzman, Gold, Andres, Arbuckle and Chaikelson, 1987), and that the positive manifold first discovered by Spearman (1904) is a consistent and surprising finding. These facts must be considered to have been *established* and research must move on to *explaining* individual differences in intelligence. This might be expressed in terms of validating the concept of intelligence. Another theme in the accounts of intelligence mentioned above is the way in which this validation might be done.

Validation of the concept of intelligence is possible by establishing the causes and the consequences of individual differences in psychometric intelligence. In a way that is similar to Sternberg's (1990) 'looking inward' and 'looking outward' classification of approaches to intelligence research, the present author conceives of attempts to validate intelligence in terms of 'looking up' to the consequences in life that are predicted by



intelligence test scores and 'looking down' to the psychological and biological bases or causes of individual differences in intelligence. 'Looking up' is not a principal concern of this thesis; Brand (1987a), Brand, Caryl, Deary, Egan and Pagliari (1991) and Jensen (1984) have provided reviews of those aspects of human performance and real-world achievement that are predicted by psychometric intelligence differences. For example, IQ has been shown to be the major psychological predictor of job success in the USA (Hunter and Hunter, 1984). However, the fact that intelligence is the major predictor should not be taken to indicate that correlations between intelligence and real-world achievements are high, in fact they are modest, but higher than correlations of life-data with other psychological variables. Indeed, some have taken the typically modest size of such correlations to indicate that IQ-type test scores do not offer much by way of predictive validity outside the realm of academic achievement (Howe, 1990).

However, this thesis is concerned with 'looking down' to intelligence in an attempt to discover its causal bases, i.e. in furthering that research which has attempted to discover which biological and psychological variables underlie individual differences in human intelligence. Previously, the 'looking down' approach has been dubbed the 'biology of intelligence' (Mackintosh, 1986; Deary, 1988a) but, because some of the approaches considered under this classification have proved to be psychologically complex, it seems to be appropriate to separate reductionist accounts into those which are clearly biological and those which are more psychological.

#### **1.4 The 'biology' of intelligence**

If individual differences in psychometric intelligence were known to be rooted in biology, then the validity of the concept of intelligence would be more assured. This approach has its own problems. To some, the biology of intelligence has been seen as a misguided exercise in genetic reductionism applied to an arbitrary and culture-biased set of puzzles (Gould, 1981; Rose, Lewontin and Kamin, 1984). Another problem is that the biology of

intelligence is very diverse; it refers to no one area of study and it is often conducted without an explicit theory of why the biological variable of interest should be related to individual differences in psychometric intelligence. These reservations notwithstanding, at least four biological approaches to human intelligence may be identified.

#### **1.4.1 Biology as a model**

First, biology may provide the model for intelligence. In Sternberg's (1985a) review of models of intelligence he included Piaget's work as an example of the biological approach to intelligence. In this sense Piaget's contribution has been important but very different to that of the psychometricians. Whereas Piaget's tests of conservation, etc. have provided mental tests with substantial *g* loadings (Jensen, 1980), and theories stressing interaction in development, his writings on the biological aspects of knowledge growth have attracted much less attention. According to Piaget (1971, 1978, 1980), knowledge accretion involves the brain in acting in a way that is analogous to the working of other organs; there is a substrate (information drawn from the environment) and there are products which are the 'post-digestive' transforms of the substrate (knowledge schemata, formed by the processes of assimilation and accommodation). Piaget's writings have a structure to their theorising that is close to that of the evolutionary epistemologists (Wuketits, 1986; Campbell, 1974; Deary, 1988b) and to those workers in artificial intelligence whose first assumptions and model constraints involve what is known about brain evolution, structure and development (Edelman and Reeke, 1982). This form of the biology of intelligence, while it might provide a metaphysical structure for psychometricians, is not of immediate concern here, because it tends to proceed in a philosophical rather than an empirical fashion (Plotkin, 1987) and, therefore, is not likely to uncover the causal bases of intelligence differences.

#### **1.4.2 Biology as genetics and race**

Second, the biology of intelligence may be seen as involving the study of genetic



contributions to individual differences in psychometric intelligence. Probably the most authoritative evidence on the heritability of psychometric intelligence has come from the Minnesota study of twins reared apart (Bouchard, Lykken, McGue, Segal and Tellegen, 1990). In this study of more than 100 sets of reared apart twins or triplets on IQ-type tests, the results indicated that about 70% of the variance in IQ test scores was attributable to genetic factors. However, the alleged fraud in this area said to have been perpetrated by Burt (Hearnshaw, 1979; Joynton, 1989) continues to make others wary of accepting the results of such approaches.

This aversion to biological reductionism is even more in evidence with the study of racial differences in intelligence, where reactions to the writings of Jensen (1969) and Eysenck (1973) have been forcibly put forward and have had considerable influence (Kamin, 1974; Gould, 1981; Rose, Lewontin and Kamin, 1984). The debate about the origins of racial differences in IQ test scores continues; Jensen (1985) argued that the one standard deviation difference in IQ scores between American blacks and whites has its source in the g loadings of the tests which, in turn, have a genetic basis. On the other hand, Mackintosh's (1986) research in this area has led him to suggest that it is social differences between groups that lead to their having differences in psychometric intelligence.

Whatever the final reckoning of the contribution of genetics to the within and between group differences in psychometric intelligence, it may be said that to know that the genetic contribution is 30, 50 or 70% does not add much to a psychological understanding of the causes of intelligence differences. Such an understanding is more likely to come from the application of modern molecular genetic approaches, which promise to unravel the actual gene products that are involved in the production of individual differences (Plomin and Rende, 1991).

### **1.4.3 Biology as neurobiology**

This endeavour involves searching for biological, often biochemical correlates of intelligence test scores. Some research relating mental ability differences in groups such as Down syndrome and Alzheimer disease patients has suggested that there are significant correlations between mental ability test scores and individual differences in biochemical processes involved in clearing the products of oxidation from neurons (Weiss, 1984, 1986; Sinet, Lejeune and Jerome, 1979; Inouye, Park and Asaka, 1984). A few reports have appeared which relate individual differences in ability test scores to indices of brain metabolism using positron emission tomography (PET; Chase, Fedio, Foster, Brooks, DiChiro and Mansi, 1984; Haier, Siegel, Neuchterlein, Hazlett, Wu, Paek, Browning and Buchsbaum, 1987). However, the first of these two studies indicated that, in Alzheimer disease patients and in the elderly, higher mental ability is associated with higher levels of glucose metabolism, whereas the latter report suggested that, in younger subjects, higher mental ability is associated with lower levels of glucose metabolism (Deary, 1988c). Both of the above PET studies used small numbers of subjects but, as the technique becomes less expensive and less cumbersome, and as the availability of more specific neurotransmitter ligands increases, it might be expected that PET scanning will contribute to an understanding of the brain metabolic processes contributing to individual differences in intelligence. Other examples of the neurobiological approach are the preliminary findings relating individual differences in nerve conduction velocity to IQ scores (Barrett, Daum and Eysenck, 1990), and the continuing debate about whether brain size relates to IQ (e.g. Rushton, 1991).

### **1.4.4 Biology as electroencephalography (EEG) and brain evoked potentials (EPs)**

This is a large research area, which demands considerable technical knowledge from the reader, and which is published in diverse journals. It has been the subject of a recent extensive review by Deary and Caryl (in press), who reviewed the entire literature in this

area published up to the middle of 1991. They concluded that various EEG and EP measures did correlate significantly with psychometric ability test scores, and that significant correlations had been found in studies of normal adults, not just in studies involving mentally handicapped subjects and children. However, the factors leading to such associations were not well understood and, therefore, despite a number of significant associations, little about the brain processes underlying IQ test score variance had been revealed by such research. Whereas studies of ongoing EEG had been relatively successful in obtaining relationships with intelligence, standard EP component latency studies had done less well, though EP studies emphasising waveform complexity and component amplitudes had been more successful.

Deary and Caryl (in press) found one particularly promising strain of research. Several independent studies had discovered that EP differences associated with early stimulus processing were related to mental ability test scores (Rhodes, Dustman and Beck, 1969; Blinkhorn and Hendrickson, 1982; Haier, Robinson, Braden and Williams, 1983; Stough, Nettelbeck and Cooper, 1990; Zhang, Caryl and Deary, 1989a, 1989b; Gilbert, Johnson, Gilbert and McColloch, 1991). In addition, Deary and Caryl (in press) argued that an atheoretical, descriptive approach was warranted as the way forward in this research. Much of the theorising in this field has been unhelpful and esoteric (Hendrickson and Hendrickson, 1980; Liberson, 1989; Weinberg, 1969; Weiss, 1987; Giannitrapani, 1971), and an approach that detailed the temporal and topographic associations between brain electrical potentials and particular mental abilities is to be preferred to one which attempts to make great theoretical leaps from intelligence test performance to detailed brain mechanisms.

In summary, the promise from molecular genetics, PET scanning and EEG and EP approaches to the understanding of individual differences in intelligence is great, but their contribution to date has been limited. One probable reason for this state of affairs is the

fact that biological variables are at a very different explanatory level when compared with cognitive ability test performance. Therefore, an intermediate level of explanation might be more successful, i.e. one which attempts to understand individual differences in ability test performance in terms of basic psychological processes which might, in turn, have tractable biological origins. Such an approach was considered by Howe (1988) and Sternberg (1988) to have great potential in 'unpacking' the bases of individual differences in intelligence, and thereby offering validation to the concept of intelligence.

### **1.5 Psychological 'components' underlying psychometric test performance**

If attempts to explain individual differences in psychometric test performance in terms of biological variables have not been successful, then a more appropriate approach might involve the uncovering of those basic psychological abilities or ability components which underlie test performance. Such an interest began with Galton (1883), who envisaged mental ability differences in terms of lower level differences in sensory discrimination. This approach to intelligence suffered a decline of interest with the growth of the atheoretical mental testing movement, and has recently seen a revival, with several lines of research attempting to uncover those mental components that underlie psychometric intelligence differences (Jensen, 1985b). However, as the approach gathered pace in the 1970s, with the arrival of experimental psychology-based tests produced by modern cognitive psychology, Hunt (1980) warned that,

A naive, but common, way of studying individual differences in cognition is to establish a statistical relationship between performance on psychometrically defined intelligence tests and performance on more theoretically defined laboratory tasks.

...While these studies should aid in advancing our understanding of the relationship between psychometric and information processing theories, the results to date do not indicate that they will produce a major breach in the 0.3 barrier. They may push it back to 0.4, but the search for a 'true' single information-processing function underlying intelligence is likely to be as successful as the search for the Holy Grail.

Several helpful reviews of various approaches that fall into this category of research are

provided by Vernon's (1987) book on information processing approaches to intelligence. Before presenting a summary of the results obtained using inspection time, two of the 'componential' approaches to psychometric intelligence, those that have been arguably the most widely-researched and successful to date, will be discussed to illustrate the success or otherwise of these endeavours. The second of these conforms to Hunt's (1980) model quoted above and will examine the research on the Hick reaction time paradigm. However, the first approach to be examined took a somewhat different line.

### **1.5.1 The componential approach of R.J. Sternberg**

Sternberg (1985b) dubbed the approach referred to by Hunt (1980) above as the 'cognitive correlates' approach to intelligence. In other words, theoretically derived cognitive test variables were correlated to test performance scores in an effort to discover which of these predicted individual differences in IQ test scores. R.J. Sternberg's own, very different approach to understanding the psychological bases of intelligence test performance was dubbed by him the 'cognitive components' approach to intelligence (Sternberg, 1985b). Sternberg (1977) attempted to discover which components underlie reasoning items involving analogies, and how they interrelate. He did this, not by correlating reasoning performance with performances on other information processing tests, but by taking apart the analogy items themselves.

Sternberg (1977) tested various models to account for subjects' performances on analogical reasoning test items. He assumed that the successful completion of such analogy items required the following components of performance: *encoding* of the first and second analogy terms; *inferring* a relationship between these items; encoding a third term and then *mapping* relations between the first and third terms; and *applying* the inferred relation to the third term in order to choose the correct answer option. Kline (1991b) criticised Sternberg's approach, stating that these so-called components are necessarily correct, i.e. analogy items could not be solved without encoding, inference, mapping and

application. According to Kline (1991b) this made the components correct *a priori* and not a matter for empirical demonstration. However, Kline (1991b) appeared to have missed the point of Sternberg's original investigation, because Sternberg did not claim to have demonstrated the existence of the components empirically. In fact, Sternberg (1977) devised a method which purported to separate the overall item completion times into the times taken by each component and, further, he attempted to discover whether each component was applied exhaustively to the analogy items or whether processing done by each component was halted when a likely solution was found.

Sternberg (1977) devised an ingenious method for presenting parts of each analogy in a tachistoscope, prior to presenting the whole analogy. This allowed him to estimate, using simultaneous equations and a 'subtraction model' of analogy response times, the time taken by each component. For instance, if a subject was shown only the first term of the analogy, e.g. 'Lincoln', he or she would be able to encode this item. Sternberg would allow unlimited time for this processing to take place. The whole analogy would then be shown and the subject asked to respond with the correct answer as quickly as possible. Sternberg reasoned that the response time for the whole analogy in this case would be faster than the response time to an analogy originally presented in complete form, and that the difference between the two response times would be the time taken to encode the first term. If a subject was shown the first two terms of an analogy problem, say 'Lincoln:Washington', he or she would be able to encode both of these items and then infer a relationship between them, before seeing the whole analogy and responding with the correct answer. Seeing the first three terms of the analogy, e.g. 'Lincoln:Washington::5:?' before the entire question was presented would remove the time required for three encoding operations and the inference and mapping operations from the response time to the whole analogy. Thus did Sternberg attempt to find out the proportion of response times taken by the different 'components' of analogical reasoning. (The correct answer to the above analogy is '1', i.e. the corresponding dollar value of the note



that contains the named president's portrait.)

After testing four different models, Sternberg concluded that the application and mapping components were self-terminating and that inference might be exhaustive. However, the difference between the fits of the different models was often very small, implying that there was often little to choose between them. Sternberg (1977) showed that over 50% of the solution time in verbal analogies was taken up by encoding of the terms, and that the response component took a substantial proportion of time. Individuals pre-selected for their high reasoning ability appeared to fit the models better and had faster responding times, but were slower at encoding. The multiple correlation between reasoning scores and component latencies was over 0.7.

Mullholland, Pellegrino and Glaser (1980) replicated some of these findings and demonstrated regularities in the increases in response times as additional stimulus elements and transformations were added. Perhaps the largest single body of research on Sternberg's componential method was reported by Sternberg and Gardner (1983), where the method was applied to analogy, series completion and classification items each in verbal, picture and geometric forms. This study showed that the components used in Sternberg's (1977) original study of analogy items were not reliably identified as significant parameters from different tasks, and that some components emerged from some tasks but not others. The components of inference, mapping and application, whose characteristics appeared to be the main discovery of this method (Sternberg, 1977), could not be separated reliably in the study by Sternberg and Gardner (1983), and they had to be combined into a 'reasoning' component (sic). When component latencies were averaged across task content, in order to compare the three types of task, the mean intercorrelation among components with the same name in different types of task was 0.32, whereas the mean correlation among component latencies for components with different names was 0.24. Similar results were obtained when the component latencies were collapsed across

task types in order to compare different task contents. Sternberg and Gardner (1983) argued that this afforded some convergent and discriminant validity for their componential theory, though Whitely (1980) had found little evidence for cross-task correlation of corresponding components.

The poor inter-component correlations of Sternberg and Gardner (1983) notwithstanding, there were other considerations that limited the conclusions that might be derived from this approach. They analysed the results of four components: encoding, reasoning, justification and comparison. First, encoding often had a poor fit to their data. Second, the reasoning 'component' was a composite that had locked within it most of the components that were earlier reckoned to be of interest. Third, the justification component did not exist in some tasks. Fourth, the comparison component was another composite, not found in the original list. Alderton, Goldman and Pellegrino (1985) found that, even given the correct processing of analogy items, high ability individuals tended to be helped to the correct answer by perusing the answer options, whereas lower ability individuals were often distracted away from the correct answer to an incorrect alternative. In this study, same name processes across tasks tended to correlate at about the same level as different level processes.

In summary, the knowledge obtained about the bases of intelligence differences obtained from the Sternberg subtraction approach to the dissection of IQ-type test items has not lived up to its early promise. Rarely were competing models of component function tested competitively to estimate whether one was significantly better than another. So-called independent processes correlated significantly in speed as well as accuracy, making a *g*-type explanation of the Sternberg results possible. Same-label processes correlated at only 0.3 across tasks, and sometimes had zero correlation. Although the chopping up of tasks appeared intuitively correct at first, as Kline (1991b) recognised, this dissection appeared more arbitrary as follow-up studies failed to demonstrate the existence of the said



components as significant parameters in the models of item solution. It was not clear how generalisable Sternberg's components were intended to be. As stated by Hunt (1980), the interest in the components underlying IQ test items lies in the possibility that such components will also be important in real world tasks; it was not clear that Sternberg's components had any life outside the IQ-type items themselves or, indeed, how the independent existence of such processing components was to be demonstrated. Whitely (1980) emphasised that Sternberg's approach was not successful at modelling response accuracy to analogy test items and that, "...no explicit mechanism is postulated to relate subject differences in the components to subject differences in total response time".

### **1.5.2 Hick reaction time and cognitive ability**

If the attempt to split IQ test items into basic processing components has not met to date with conspicuous success, what of the attempts that begin with components from theoretically well-understood cognitive tasks and attempt to correlate these to scores on psychometric ability tests? For as long as psychology has been an empirical endeavour, some form of reaction time has attracted the attention of those who fancied that they could understand intelligence in terms of speed of information processing (Galton, 1883; Wissler, 1901).

As early as 1933 Beck had reviewed over 30 studies which had examined the relationship between reaction time and tests of mental abilities. In all instances, negative correlations should be taken to indicate that brighter subjects had faster reaction times. Briefly, Beck found: 14 intelligence correlations with simple and discriminative reaction times ranging from 0.32 to -0.90 with a median of -0.16 (the high result was the surprising report by Peak and Boring in 1926 after testing 5 senior students); 14 correlations of serial reaction time and intelligence with a range of 0.03 to -0.53 and a median of -0.18; 5 correlations of intelligence and speed of reading giving a range of -0.14 to -0.32, median -0.30; 6 correlations looking at intelligence and speed in serial verbal tasks with a range of 0.06 to

-0.23, median -0.12; and 6 correlations of intelligence with speed of reflex response latency ranging from 0.08 to -0.24, median -0.06. The review by Beck was incomplete. It did not include the Travis and Hunter (1928) report of a correlation of -0.87 between intelligence as measured by the Otis and Iowa exams and the time between tapping the patellar tendon and the arrival of the motor nerve impulse at the quadriceps femoris muscle.

In recent years the focus of intelligence researchers has been on reaction time paradigms with an arguably better theoretical underpinning. Sternberg's (1966, 1969) rapid memory scanning reaction time procedure has attracted some research attention in the field of intelligence (Chiang and Atkinson, 1976; Puckett and Kausler, 1984; Todman and Gibb, 1985; Jensen, 1987a; Deary, Langan, Graham, Hepburn and Frier, in press) but, although psychometric test scores tended to correlate significantly with reaction times from this paradigm, the theoretical difficulties with the Sternberg test, especially its slope, have led to little theoretical advancement about the bases of intelligence in terms of information processing mechanisms. Correlations between cognitive ability test scores and Sternberg rapid memory scanning test parameters have tended to be with overall reaction times or with the intercept (Todman and Gibb, 1985), but not with the hypothetical components which the slope was originally thought to comprise (Jensen, 1987a; Deary, Langan, Graham, Hepburn and Frier, in press).

The reaction time procedure which has attracted most attention in intelligence research has been the so-called Hick paradigm, and a very thorough review of this research was undertaken by Jensen (1987b). Blank (1934) noted that reaction time (RT) increased as a linear function of the logarithm of the number of stimulus responses, and Hick (1952) expressed this in terms of information processing theory, where a 'bit' of information was equal to a binary choice. The potential importance of a relationship between individual differences in the increase in RT as stimulus uncertainty increased and individual differences in intelligence was raised by Roth (1964; and see Eysenck, 1967), who found

a correlation of -0.39 between the slope of the Hick paradigm RT regression line and scores on a psychometric test of intelligence, but a zero correlation between the intercept from the Hick RT paradigm and ability test scores. In other words, brighter subjects' RT slopes increased less steeply as the number of response choices increased on the Hick RT task. Therefore, it seemed that individual differences in intelligence might be related to individual differences in the 'rate of gain of information' that were thought to be indexed by the slope of the Hick RT paradigm.

Jensen's (1987b) review of research in intelligence using the Hick paradigm examined 33 study samples from 27 studies, involving a total of 2,317 subjects. Twenty-one of the samples involved Jensen as author or co-author, approximately half of the samples included college or university students, and others included above-average ability children. The average fit of the RT data to Hick's law was 0.995 for decision times, while movement times tended not to fit an increasing slope. Decision time (DT) will be used here to refer to the time taken to lift the finger from the 'home' button on the reaction time apparatus after the stimulus light has been lit (this is called 'reaction time' by Jensen, which is confusing, since the overall response time has the same acronym), whereas movement time (MT) will be used to refer to the time taken to press the response button corresponding to the appropriate stimulus light, after the finger has been lifted from the 'home' button. In the Jensen (1987b) review, the N-weighted mean correlations across studies between parameters derived from the Hick paradigm and tests of mental ability were as follows (with correlations corrected for unreliability and restricted range given in parentheses): -0.31 (-0.32) for overall DT mean; -0.18 (-0.25) for DT intercept; -0.18 (-0.28) for DT slope; -0.32 (-0.48) for DT variability; -0.29 (-0.30) for MT mean; and -0.02 (-0.02) for MT variability. Negative signs preceding the correlations indicate that brighter subjects were faster and less variable on the Hick RT task parameters.

Therefore, there were generally low correlations in the expected direction with various

Hick RT procedure parameters. DT slope, the parameter originally thought to tap individual differences basic to psychometric intelligence, had no exclusive correlation with ability scores, and the correlations with slope were not the highest. Even MT variability was related to differences in cognitive ability. Jensen (1987b) argued also that there should be higher correlations between estimates of intelligence and the DTs associated with greater degrees of stimulus uncertainty. He offered a summary of evidence from 15 independent groups to show that the correlations between cognitive ability estimates and DTs for 0, 1, 2 and 3 'bits' of information were -0.19, -0.21, -0.24 and -0.26, respectively. Therefore, the differences between these correlations were very small, as was the absolute size of the correlations, but they ran in the expected direction, given the hypothesis that the higher ability person had a greater rate of gain of information. However, several technically adequate studies did not find this relationship.

Jensen (1987b) concluded that correlations between Hick parameters and IQ scores were based upon some general speed and/or efficiency reflected in most aspects of performance in the Hick paradigm. Jensen even considered the hypothesis that the general factor extracted from a battery of RT tasks might be the same as the *g* factor extracted from a battery of psychometric ability tests. In addition, Jensen attempted to refute explanations of the Hick RT-IQ correlations based upon the following hypotheses: that there was a common test-taking factor; that high ability subjects operated a speed-accuracy trade-off; and that high ability subjects were characterised by their high motivation or arousal.

A series of problems associated with the Jensen apparatus<sup>p</sup> used to estimate Hick RT parameters was raised by Longstreth (1984) who suggested that there were order effects, visual attention effects and response biases associated with the apparatus. Longstreth (1984) also demonstrated that the DT slope-IQ correlation, and the association between DT complexity and IQ did not hold when groups of normal subjects were considered. Jensen and Vernon (1986) and Jensen (1987b) answered much of the criticism raised by

Longstreth (1984). However, perhaps the most worrying criticism by Longstreth was that, in the typical Hick RT procedure, the trials with the small degrees of stimulus uncertainty occur first, with the trials associated with the greater degrees of stimulus uncertainty following in increasing order. Therefore, it might be the case that the higher IQ individuals learn the RT task faster and therefore, the correlation between Hick RT slope and IQ might come about because the learning effect is confounded with stimulus uncertainty.

Widaman and Carlson (1989) administered the Hick RT procedure in the standard ascending manner and added conditions where the degrees of stimulus uncertainty were met in descending and random orders. As predicted by the learning effect hypothesis, the slope for RTs was steepest in the ascending condition, intermediate in the random condition and flattest in the descending condition. More importantly, Widaman and Carlson (1989) found that the correlations with IQ tended to go in the direction that allied IQ with those who become faster with practice; i.e. IQ appeared to be related more to the rate of RT increase with practice than with RT components *per se*. Therefore, intelligence appeared to be related to individual differences in the rate of automatising of a new task, as articulated for other tasks by Ackerman (1990). Widaman and Carlson (1989) suggested that correlations between Hick RT parameters and cognitive ability might be ephemeral.

In summary, the state of research with the Hick RT paradigm is not dissimilar to that found with the Sternberg componential approach. There are some small significant correlations between cognitive ability test scores and Hick RT parameters, but they are not sufficiently tied to theoretically important aspects to offer much hope that the Hick paradigm will lead to enlightenment concerning the sources of individual differences in intelligence. Additionally, there remains the possibility that Hick RT-IQ correlations are due to unexpected aspects of Hick RT performance, such as rate of improvement on the task.



## **1.6 Inspection time**

This thesis will focus on research which has attempted to relate individual differences in inspection time to individual differences in cognitive ability. Inspection time (IT; Vickers, Nettelbeck and Willson, 1972) has been defined as, "the minimum exposure time needed for observers reliably to identify a highly evident feature of a stimulus display" (Levy, in press). IT has the immediate advantage that it does not have the motor response component variance of any of the reaction time procedures. Also, it is rooted in psychophysical theory and, therefore, has a relatively well-founded theoretical basis. While the primary concern of this thesis will be to study the relationships between an auditory form of inspection time and IQ-type test scores, the main body of research in this area concerns visual inspection time, and a brief review of the main results will be presented here as an introduction to the field. Studies involving auditory information processing and ability test scores will be considered in detail in Chapter 2.

### **1.6.1 Historical precedents of inspection time**

The discovery that there were individual differences in the stimulus duration needed to perceive a stimulus and make a discriminative judgment about its features to a given level of accuracy was made by McKeen Cattell while he was carrying out investigations for his Ph.D. in Leipzig (Deary, 1986). Cattell reported his early investigations in two papers in 1886 (Cattell, 1886a,b). Cattell was able to assemble a small review of the ideas and reports of earlier investigators who had had similar ideas. These reports stretched back to the 1830s, but it was with the arrival of the accurate tachistoscope and chronometer that Cattell was able to give the idea a firm basis with his empirical findings. Cattell's concern was that the reaction time was too motor and that the individual elements of the reaction time would reveal more about the timing of mental events. Cattell (1886a,b) devised his "perception time" as the time needed by a subject to make a single discrimination correctly; i.e., the time needed *to see*, freed from the constraint of having to respond quickly, a

stimulus and be able to report accurately that it was one thing and not another. For example, Cattell had his subjects discriminate various colours from a standard grey, and he found that there were reliable individual differences across people and across colours in the times required in order to make these discriminations (correlations for these results were computed by Deary, 1986).

Despite the discouragement of the influential Professor Wundt, who was far from sympathetic to the investigation of individual differences (Cattell and Farrand, 1896), Cattell was able to discover that his few subjects maintained their individual differences in "perception time" when they moved from colours to letters and then to words. In his reports, Cattell (1886a,b) also anticipated the effects of backward masking: he briefly alluded to the fact that what followed the stimulus had a significant effect on the perception time. Also, he anticipated the phenomenon of 'chunking' in immediate memory, by emphasising that the perception time for a single letter and a word were about the same. And he discovered that the number of ideas that can be held in consciousness was about 5 (within Miller's magic number seven plus or minus two). The possible relationship between perception time and more general mental ability was only hinted at in the experiments that Cattell (1886b) performed on his subjects,

The individual difference is a matter of special interest. B out of 40 trials read correctly 5 times a card containing 7 numbers and could always read 5 numbers correctly. He could grasp 6 letters, four disconnected words, or a sentence of seven words, whereas others could grasp but 3 letters, 2 words, or a sentence of four words.

The latter numbers are the limits for one of the four students experimented on, and for the two women, one an educated young lady, the other the wife of a mechanic. The limit for a boy of nine years old was somewhat higher. I tried to make the determinations on two rather obtuse porters, but their consciousness did not seem able to take up at all such delicate impression. They required three times as long as educated people to read a word.

It is odd, on the face of it, that this elegant basic measure - the stimulus duration needed in order to see or, perhaps, to hear a stimulus accurately - was never included in the

batteries of 'simple' measures when they were administered to hundreds of American college students or English schoolchildren (Wissler, 1901; Spearman, 1904; Burt 1909-10). The later researchers of the 1890s and the early 1900s seemed to prefer reaction times or simple sensory discrimination measures, as will be seen in Chapter 3. Wissler's (1901) study used a test battery including reaction time and many anthropometric body measurements, and it was the result of almost ten years of effort by McKeen Cattell after returning to Columbia. The development of Cattell's test battery away from his emphasis on "perception time" might have stemmed from the lack of encouragement he got from Wundt and the influence that Francis Galton had on his ideas when he spent some time at Cambridge. The account of Cattell's psychological and anthropometric battery as it moved from the famous "Mental Tests" account in 1890 to the Farrand and Cattell account in 1896 made it appear to be more and more of a psycho-anthropometric blunderbuss, largely deprived of psychological theory and of Cattell's (1886a,b) earlier emphasis on the "elements" of the reaction time, including "perception time". Spearman's (1904) amusing and denigratory account of the testing that Wissler (1901; supervised by Cattell) attempted was the bitter fruit of this labour.

After Cattell (1890; Wissler, 1901) had allowed "perception time" to slip from his battery of mental tests, it is interesting to note that, although speed of perception did not figure largely in the debates of simple correlates of intelligence, results reporting associations between speed of visual processing and cognitive ability differences still appeared from time to time in the literature. Griffing (1895-6), testing groups of schoolchildren, exposed black letters on a white board using a tachistoscope to give accurate 100ms exposures. Griffing took heed of Cattell's results on letter legibility and estimated the average number of letters that each child could extract from a single exposure. Griffing (1895-6) discovered,

[that] 9 of those rated A for mental capacity by the teachers, on an A,B,C basis, had somewhat higher averages than others and out of the twelve best observers (four from each age group) eight were rated A and but one C.



Thus, the amount of information extracted from a single brief exposure of a stimulus was related to mental ability as estimated by the teachers. From Griffing's (1895-6) class data the correlation between mental age and the letters seen in a single exposure was almost unity. Of course, this involves the potentially confounding factors of chronological age and reading experience and must be seen as no more than indicative.

Burt was the next investigator to find an arguably similar relationship. The Spot Pattern Test used in his first empirical study (Burt, 1909-10) involved the subjects being given a series of 25ms exposures of dot patterns arranged in a 5 x 5 matrix. The subjects were given sufficient exposures until they reproduced the pattern exactly. This would involve some memory also, but the manipulated variable was the accumulated exposure time, and this correlated at above 0.7 with teachers' estimates of intelligence. Several factors seem to have prevented Burt from following up this potentially important result: he interpreted the test as a high order test of accumulated apperception (following McDougall's discussion of the test in his *Physiological Psychology* (1905)); Burt disliked using apparatus with children because he thought it inspired "needless apprehensions"; and Burt had found a single test that had correlated even more highly with estimated intelligence in his study - the ability to dot irregular circles in a moving tape.

The idea that those who were quick on the uptake were the more intelligent appeared independently much later. Livson and Krech (1956) tested the idea that so-called "cortical conductivity" was the basis of intelligent functioning. They tested 22 college sophomores (10 males, 12 females) for their ability to reproduce dot patterns from brief tachistoscopic exposures and related individual differences on this ability to scores on the Wechsler Vocabulary Scale. The result was a correlation in the expected direction of 0.54 ( $p < .01$ ). That such a simple task should correlate with a measure of vocabulary needed some explanation (the authors were unaware that it had already been anticipated by Burt

(1909-10)),

What has this to do with intelligence? We are suggesting that intelligence test performance is a behavioural manifestation of cortical conductivity as it mediates the efficiency of cortical functioning. This postulation of a general intelligence factor is immediately reminiscent of Spearman's conception of "g." However, by linking "g" with a neurological model, we hope to rescue this generalisation of the well-established communality in "cognitive" performance from its telling denouncement as a "mathematical artifact".

The above quotation shows that Livson and Krech (1956) had similar concerns to those that are central to this thesis, i.e. that psychometric intelligence may be validated by linking individual differences on IQ-type test scores to fundamental psychological processes or components.

Apart from these occasional findings, those who sought to relate individual differences in cognitive ability to some form of mental speed have tended to focus upon reaction times, or EEG or EP correlates of intelligence (Jensen, 1982; Deary and Caryl, in press). However, the concept of inspection time (IT), as conceived and developed by Vickers and his colleagues (Vickers, Nettelbeck and Willson, 1972; Vickers and Smith, 1986) focussed the attention of researchers in the field of human intelligence once more on the speed of information processing using briefly presented visual material.

### **1.6.2 The theoretical basis of inspection time**

The idea of an 'inspection time' (IT) has a history in the idea of the 'perceptual moment', i.e. the notion that perception operated in a quantal fashion such that a stimulus must be present from the beginning of a perceptual 'sampling period' for a sufficient duration in order to be discriminable (e.g. Shallice, 1964). Vickers, Nettelbeck and Willson (1972) developed the accumulator model of perception of Vickers (1970) to develop the IT index as,

...the time required by a S to make a single observation or inspection of the sensory input on which a discrimination of relative magnitude is based.

Thus, Vickers and colleagues (Vickers, Nettelbeck and Willson, 1972; Vickers and Smith, 1986) suggested that the amount of time required by a subject in order to make a given discrimination of relative magnitude under given conditions might represent a stable characteristic of an individual's perceptual performance.

The typical visual IT task involved a subject in discriminating which of two briefly presented parallel, vertical lines of markedly different lengths was longer (Vickers, Nettelbeck and Willson, 1972). Typically, the lines were presented in a tachistoscope and backward masked with a pattern mask to prevent further stimulus processing (Turvey, 1973; Breitmeyer, 1984). The difference in the lengths of the lines was set to be so large that the visual angle they subtended was sufficient to make the discrimination easy, i.e. affording perfect performance at longer presentation times (Vickers, Nettelbeck and Willson, 1972). Responses in IT tasks were typically *unsped*ed, with accuracy being emphasised over speed of responding. Therefore, the task was set up to be a relatively pure test of speed of visual processing, freed from requirements to react quickly or to make difficult spatial discriminations.

Vickers' (1979) accumulator model of discrimination hypothesised that the subject who was attempting to make a discrimination was sampling from the stimulus against a background of noise. Where the discrimination was a two-choice task, the accumulator model stated that evidence was accumulated in two 'counters' and, once the evidence in one counter had passed a threshold, a response was given. Although there are other conceptualisations of the type of discrimination which is involved in IT tasks (Link and Heath, 1975), the accumulator model appears to offer a credible account of the perceptual and decision-making processes involved in the task (Vickers and Smith, 1985, 1986). Recently, Vickers has attempted to expand the range of tasks which are based on the accumulator model, with some success (Foreman, 1991).

### 1.6.3 Inspection time-cognitive ability associations

The idea that IT might be a basic limitation to general cognitive performance, especially among the mentally handicapped, was conceived by Nettelbeck (Nettelbeck and Lally, 1976), and this was developed into a more general mental speed theory of intelligence by Brand (1981). Of Nettelbeck and Lally's (1976) pioneering attempt to correlate IT with cognitive ability Vickers and Smith's retrospective account (1986) stated that,

...one major strategy guiding attempts to measure the speed of mental functioning has been to isolate some process sufficiently elementary to be relatively immune from higher cognitive activities or by motivational and social factors. In its focus on a simple, component process, likely to play a limiting role in most (if not all) more complex processes, this strategy resembles the employment of standard algorithms as benchmark tests of processing speed of a digital computer. In the field of human information-processing research, it has been argued, an analogous measure of speed is provided by an inspection time (IT) index...

If IT does measure the time required to make a single observation of the sensory input, then such a quantity seems likely to operate as a basic factor limiting perceptual and cognitive performance in general. In agreement with this, and following the suggestion by Savage (1970) that differences in intellectual ability might be attributable largely to the speed with which 'some kind of search or scanning mechanisms operate,' Nettelbeck and Lally (1976) and Lally and Nettelbeck (1977) examined IT in two samples of young adults, and found high correlations, in the range -0.8 to -0.9, between IT and Performance IQ as measured by the Wechsler Adult Intelligence Scale.

Therefore, as is clearly indicated above, the IT index appeared on the scene like a glint from the pillar in which the 'Holy Grail' of intelligence research was hidden (Hunt, 1980); IT might be a fundamental psychological processing index, uncontaminated by higher level psychological and social variables; might be ubiquitous, infusing most cognitive operations with its variance; and might offer a 'benchmark test' of human mental speed, some form of which had long been thought to be basic to individual differences in IQ (Berger, 1982). Here, it seemed, was a test that might validate the notion of intelligence by associating individual differences in IQ-type test scores to a fundamental information processing index. The findings of Nettelbeck and Lally (1976) and Lally and Nettelbeck (1977) began this line of reasoning (Brand, 1981), though their subject samples had exaggerated ranges of IQ owing to the inclusion of mentally handicapped subjects.

By 1982 Brand and Deary had undertaken a review of nine studies which had correlated IT with cognitive ability test scores. Three of these studies were the work of Nettelbeck and colleagues at the University of Adelaide, and five were undergraduate theses supervised by Brand. Brand and Deary (1982) identified five 'modal' studies (Nettelbeck and Lally, 1976; Anderson, 1977; Lally and Nettelbeck, 1977; Grieve, 1979; Deary, 1980) which had the following characteristics: the inclusion of young adult subjects; a range of IQs around 100; an IT task involving the comparison of the lengths of two lines; and the inclusion of non-verbal or 'culture-fair' measures of *g*. In these five studies the median correlation between IQ and IT was -0.8, and the results appeared to indicate a relation between IT and general intelligence rather than a particular association of IT with performance IQ, as had been suggested by Nettelbeck and Lally (1976). This high level of correlation led Brand and Deary (1982) to suggest that one of the practical applications of the IT-IQ correlation was that,

It should be possible to test fluid intelligence in a way that transparently fair to people of varying socio-economic, psychopathological, ethnic, national and racial groups

However, Nettelbeck (1982), after reviewing those studies which had focussed upon subject samples with average or above average ability, suggested that this was an incautious conclusion, which was based upon high correlations obtained from subject samples which had very large IQ ranges and included mentally handicapped subjects. Nettelbeck (1982) concluded that,

The results of our studies, however, do not support the suggestion that such a speed factor could account for more than some small part of intelligence.

There is little to be gained by discussing the results of every study which has examined the relation between IT and IQ-type test scores. There have been a number of qualitative reviews of the research (Lubin and Fernandez, 1986; Vernon, 1986; Nettelbeck, 1987;

Juhel, 1991) and one meta-analysis of the IT-IQ research (Kranzler and Jensen, 1989). There is a consensus to the effect that the earlier studies' high correlations were, as Brand and Deary (1982) had themselves pointed out, probably a result of large IQ ranges and the inclusion of some mentally handicapped subjects in some subject samples. Nettelbeck (1987) reviewed 16 sets of results, comprising 529 IT estimates obtained from 439 subjects, which had correlated IQ-type scores and IT measures involving visual or auditory stimuli (see Chapter 2), and found that the average uncorrected correlation among young non-retarded adults was -0.35. Many studies included undergraduates and, therefore, when this correlation was corrected for restriction of range,

"The best available estimate of the strength of the association between IT and general ability across the full range of IQ is about -.5" (Nettelbeck, 1987).

One major theme in the review article by Nettelbeck (1987) and in the meta-analysis by Kranzler and Jensen (1989) was the possibility that IT might be more closely related to either Performance or Verbal IQ. After reviewing 12 studies, Nettelbeck (1987) concluded that,

"...the question of an IT-verbal ability relationship remains unresolved, with some evidence consistent with this approach and some not".

Nettelbeck (1987) found stronger evidence for a reliable association between Performance IQ and IT. However, of the nine studies which Nettelbeck (1987) reviewed in forming this conclusion, only one contained a sample of normal adults; the others were composed of samples of university students or mentally handicapped adults. In the meta-analysis by Kranzler and Jensen (1989), the uncorrected correlations between IT and Verbal and Performance IQs in adult, non-retarded samples were, respectively, -0.18 and -0.45, confirming Nettelbeck's (1987) judgment that the IT-verbal ability association was less well established than the IT-performance ability association in adult samples. No single data set from any one study was sufficient to decide the issue unequivocally.



Juhel (1991) concurred with the view that there exists a modest association between IT and cognitive ability, and that the association is of theoretical interest, but emphasised that the original view of IT as a 'simple' speed of processing index had not gone unchallenged. Some visual IT tasks proved prone to cognitive strategies, such as apparent movement (Egan 1986, 1991; Mackenzie and Bingham, 1985; Mackenzie and Cumming, 1986). Moreover, it has been reported that IT estimates obtained using the same stimuli could be varied by increasing or decreasing the 'cognitive loading' on the discrimination question (Mackenzie, Molloy, Martin, Lovegrove and McNicol, 1991). Nettelbeck (1987) suggested that IT, as measured in many studies, might be indexing efficiency of "early central stages of perception" rather than solely speed of apprehension.

In summary, it is suggested here that IT research offers a better opportunity for understanding a modest amount of the variance underlying individual differences in cognitive ability than do other most prominent 'mental speed' approaches, such as Sternberg's componential method and the Hick RT paradigm. However, the IT-cognitive ability association requires two separate efforts. First, it requires to be *established*, i.e. its existence must be demonstrated to be reliable and the strength of the association should be reckoned. Although the suggestion by Mackintosh (1981) and Nettelbeck (1987), that a large scale study of IT involving subjects with a normal distribution of IQ should be undertaken, has yet to be realised, there appears to be sufficient evidence to conclude tentatively that the IT-cognitive ability association has been replicated enough times to be considered established. Further, the strength of the association has been reckoned at a correlation of about 0.35 in extant studies, correcting to about 0.5 when the attenuated range of cognitive ability of most samples is taken into account (Nettelbeck, 1987; Kranzler and Jensen, 1989; Juhel, 1991).

Second, the IT-cognitive ability association requires to be *explained*, i.e. the relationship



itself must become the object of research in an effort to understand *why* the relationship exists after it has been demonstrated *that* it exists. These two research efforts are important, but are frequently mundane and frustrating, respectively. It is mundane merely to replicate or fail to replicate a relation between two variables, but replication of psychological phenomena is necessary in an area where results are frequently evanescent. It is frustrating to attempt to explain a relationship when the research is likely to proceed up many empirical and theoretical dead-ends, and where the theoretical concepts involved are often difficult to operationalise (e.g. see the description of strategic accounts of IT performance in Brand (1987b) and in Egan and Deary (in press) and the attempts by Zhang, Caryl and Deary (1989a,b) to attempt to discover the evoked potential correlates of the IT-IQ association).

One way of overcoming the concerns about the IT index which result from its visual format is to attempt to demonstrate that there exists an association between speed of information processing and cognitive ability in another modality. Therefore, this thesis will investigate whether the auditory modality may be used to set up a parallel research effort; it will attempt to address some of the problems peculiar to an IT-type test in that modality, and will attempt to replicate results addressing some of the important issues which other researchers have addressed in the visual form of the task. The work that has been done to date on speed of auditory processing and cognitive ability will be the subject of the next Chapter.

## Chapter Two

### Intelligence and auditory information processing

*There remains, in consequence, no justification for the sharp distinction between the sensory perception of qualities and the more abstract processes of thought; we shall have to assume that the operations of both the senses and the intellect are equally based on acts of classification (or reclassification) performed by the central nervous system, and that they are both part of the same continuous process by which the microcosm in the brain progressively approximates to a reproduction of the macrocosm of the external world. (Hayek, 1952).*

#### 2.1 Auditory inspection time and intelligence

##### 2.1.1 Brand and Deary (1982)

Brand and Deary (1982) devised a so-called auditory inspection time (AIT) task with the intention of indexing speed of information processing in the auditory modality in a way that was analogous to the speed of information processing in the visual modality indexed by visual inspection time tasks. It was argued by Brand and Deary (1982) that,

If, as argued above, mental speed is the basis of general intelligence, then it is difficult to imagine why this mental speed would be manifest solely in the visual modality.

The auditory discrimination task devised by Brand and Deary (1982) took the following form: two square wave tones of markedly different pitch and instantaneous onset and offset were played consecutively, and the subject was required to respond by stating the temporal order of the two tones, i.e. 'High-Low' or 'Low-High'. A subject was alerted to the initiation of each item by a spoken "Ready" cue. The subject then heard the first tone (either 770 or 880 Hz) followed by white noise in the gap between the two tones which was played for approximately 500ms. The tone frequency control switch was then

changed manually to select other frequency, and the subject then heard the second stimulus tone (880 or 770 Hz). Each tone was played for an identical length of time, with stimulus durations in the AIT test ranging from 100ms to 2.7ms. Both tones were forward and backward masked using white noise, with no stimulus-mask gap. Stimulus tones were played at 90 dB and the masking noise was played at 74 dB.

The psychophysical procedure used in the AIT test by Brand and Deary (1982) was a method of constant stimuli (descending series), i.e. the test began with stimuli of long duration and proceeded to shorter and more difficult durations. The block of trials comprised 228 items, comprising 19 stimulus durations with 12 trials at each duration. As a training criterion for the AIT task, all subjects were required to obtain 12 consecutive correct responses at a stimulus duration of 100ms. Three of 14 non-mentally handicapped subjects in Brand and Deary's (1982) study (21.4%) proved unable to meet this criterion, and they had difficulty in making the pitch discrimination reliably at even longer durations. These subjects were classified as not being able to perform the AIT task. Non-mentally handicapped subjects were tested also on a 42 item test of unspeeded pitch discrimination. The three subjects classified as being unable to perform the AIT task obtained lower scores on this test than any of the 11 non-handicapped subjects who were able to perform the AIT task (Deary, 1980). One hospitalised, mentally handicapped patient who failed to satisfy the training criterion did prove able to make the discrimination reliably at longer durations and finally achieved an AIT threshold of 160ms. AIT thresholds were determined for each subject by examining the responses to the block of trials and selecting the last block at which he or she scored 11 out of 12 trials correct. Occasional errors at relatively long durations were allowed to be 'recouped' by superior performance at briefer durations.

The 13 subjects classified as being able to perform the AIT task in the study reported by Brand and Deary (1982) included 2 hospitalised mentally handicapped subjects. One of

these subjects was assigned an AIT threshold score of 160ms, whereas the next highest AIT for any subject in the study was 20ms, and this was the score of the other handicapped subject. With the two handicapped subjects included ( $n=13$ ), the Mill Hill Vocabulary IQ mean was 102.4 (SD 21.3, range 59 to 135) and the AIT mean was 21.3ms (SD 41.6, range 6 to 160ms). When these subjects were excluded ( $n=11$ ), the Mill Hill IQ mean was 105.4 (SD 15.4, range 72 to 135) and the AIT mean was 10.0ms (SD 2.8, range 6 to 15ms). AIT correlated at -0.70 with Raven's Progressive Matrices, and at -0.66 with Mill Hill Vocabulary IQ scores when the two subjects with mental handicap were included ( $n=13$ ). However, when only non-handicapped subjects were included ( $n=11$ ), the AIT correlations with Mill Hill IQ and Raven's Matrices scores were -0.45 and -0.19, respectively (both  $p>0.1$ ). Although the two mentally handicapped subjects who were able to perform the AIT task had the highest AIT thresholds and visual IT thresholds in the sample, when only non-handicapped subjects were included the correlation between visual and auditory inspection time thresholds was -0.05 (ns). Therefore, as has been pointed out by Irwin (1984), Vernon (1986), Lubin and Fernandez (1986), Mackintosh (1986) and Juhel (1991), the correlations between inspection time and IQ-type test scores may be inflated from non-significant values to significant levels by the inclusion of a few mentally handicapped subjects. The use of such mixed subject groups should probably be avoided when studying information processing correlates of cognitive ability and, where data are gathered from handicapped and non-handicapped subjects in the same study, their data should be presented separately (e.g. Todman and Gibb, 1985).

### **2.1.2 Irwin (1984)**

Irwin (1984) criticised the study by Brand and Deary (1982) for its small and heterogeneous subject sample and stated that his study,

...attempts to replicate Brand and Deary's (1982) result for auditory inspection time. This replication seems advisable because theirs is the sole study of auditory inspection time.

...In addition to the measures of inspection time and intelligence, a conventional

test of pitch perception was administered to determine the extent to which auditory inspection time involved a component of frequency discrimination...

Irwin (1984) tested fifty 12- to 13-year-old children. In addition to pitch discrimination testing Irwin (1984) screened all of his subjects on a standard audiometric test. The latter test revealed that none of his unselected 50 schoolchildren had a loss of greater than 15 dB at any audiometric frequency between 500 Hz and 8000 Hz. Irwin's auditory inspection time task closely replicated that of Brand and Deary (1982). Square wave tone bursts with 1ms rise and fall times, played monaurally through an earphone at 90 dB, were used as stimuli. Discriminanda comprised two tones, one 770 Hz tone and one 880 Hz tone, played consecutively, 1000ms apart. Subjects were required to state which of the tones had been played first. Whenever the tones were absent white noise played at 80 dB was present, i.e. the tones were backward- and forward-masked by white noise and the gap between the two stimulus tones was filled with white noise. An adaptive psychophysical procedure was used to determine the auditory inspection time thresholds at which subjects were 71% accurate in discriminating the stimuli (Levitt, 1971).

Irwin (1984) obtained estimates of auditory inspection time threshold from 49 subjects, implying that almost all subjects were deemed by him to be able to complete the task discrimination in a satisfactory fashion, resulting in valid AIT threshold estimates. The mean AIT for the children was 195ms (SD 339.7ms), i.e. the thresholds had a very skewed distribution. The median AIT threshold for the 49 subjects was 16ms. Bearing in mind that Brand and Deary (1982) used approximately 92% thresholds and adult subjects, Irwin's study appears to have obtained thresholds for most subjects in a similar range of durations to the threshold estimates of Brand and Deary (1982). Auditory inspection time threshold estimates correlated at -0.23 ( $p < 0.05$ ) with scores on Raven's Progressive Matrices, and at -0.32 ( $p < 0.01$ ) with scores on the Mill Hill Vocabulary Scale. Kendall's tau was used as the correlation coefficient because of the skewed AIT threshold estimates. However, the results were very similar for the AIT-Raven correlation when Pearson's

product moment correlation was used, though the Mill Hill-AIT correlation increased by about 0.1 when Pearson's coefficient was determined. Irwin (1984) also computed the correlation between AIT and visual IT. The visual IT task was not the standard two-lines discrimination, but involved subjects in indicating whether a lower case 'o' or an unlauded 'u' occurred before a mask on a computer monitor screen. The AIT-visual IT correlation was 0.17 (ns, Kendall's tau).

In his discussion of results, Irwin (1984) referred to the '0.3 barrier' that is often reckoned to exist between tests of information processing ability and psychometric cognitive ability tests (Hunt, 1980; Sternberg, 1981),

The arithmetic average of the four correlations between inspection time and intelligence found in this study is -0.2279. There is no evidence, therefore, that the 0.3 barrier has been broken.

Such small correlations, Irwin (1984) reckoned, might have resulted because of brighter subjects being less anxious in approaching the AIT tasks or being able to master new tasks more quickly than less bright subjects. This view, that individual differences in inspection time and other information processing abilities arise as a *result* of rather than form a *cause* of individual differences in cognitive ability, is one that has been voiced by others (Ceci, 1990; Mackintosh, 1986; Howe, 1988), and it merits serious consideration (see Chapters 8 and 9).

The Kendall's tau coefficient value for the association between scores on the Seashore pitch discrimination test and AIT threshold estimates in Irwin's (1984) study was -0.51 ( $p < 0.01$ , Pearson's product moment correlation = -0.54), indicating that the AIT task, which had been devised to estimate *speed* of auditory information processing, was correlated moderately highly with an unspeeded test of pitch discrimination involving unmasked stimuli. Seashore pitch test scores were correlated with Mill Hill Vocabulary Scale scores at 0.37 ( $p < 0.05$ ), and with Raven scores at 0.27 ( $p < 0.1$ ). Therefore,



unspeeded pitch discrimination was significantly related to verbal ability and tended toward a significant association with non-verbal ability. When the partial correlation between AIT and Mill Hill vocabulary was calculated, controlling for Seashore pitch scores, the correlation fell to -0.29 ( $p < 0.1$ ).

Irwin (1984) added to the above results with additional evidence to indicate that the AIT task might be indexing pitch discrimination ability *per se* rather than speed of auditory information processing. He demonstrated that the energies in the fundamentals of the 880 and 770 Hz square wave tones used in the AIT task were clearly separated in frequency when the stimulus durations were relatively long (20ms or greater). However, when the same examination was carried out on stimuli of 5ms duration, a value which was close to that of the AIT threshold estimates of several subjects in the Brand and Deary (1982) study, there was less energy in the fundamentals and the energy was spread over a wider range of frequencies. Irwin (1984) concluded that,

In other words, the two fundamentals are now more physically alike and harder to tell apart. As the duration of the two waveforms is reduced, so their frequency content becomes more similar. The measure of auditory inspection time used by Brand and Deary is therefore a measure of the minimum difference in frequency that can be detected rather than a measure of processing speed.

Two observations may be made concerning Irwin's (1984) conclusion. First, it should not be forgotten that, although the correlation between AIT and Seashore pitch scores was the highest correlation in his study, and despite his demonstrating the problems that exist in presenting the AIT stimuli at very brief durations, the stimulus tones did also decrease in duration as the energies in the fundamentals became less well separated in frequency. Therefore, perhaps a more prudent conclusion would have been to hypothesise that frequency detection had been *confounded* with processing speed in the Brand and Deary (1982) study, owing to the very brief AIT thresholds obtained by most of the subjects. This weaker form of Irwin's (1984) conclusion is supported by the fact that the AIT-Mill Hill correlation reduced, but still tended toward significance, when the partial correlation



between the two variables, controlling for Seashore pitch scores, was computed.

Second, given that the AIT test did have a significant, albeit modest correlation with cognitive ability test scores, Irwin's (1984) conclusion that the AIT test indexes individual differences in frequency discrimination, and his finding that Seashore pitch test scores correlated at 0.37 ( $p < 0.05$ ) with Mill Hill vocabulary scores were remarkable. This appeared to herald a return to a belief in the existence of a significant association between sensory discrimination and measures of cognitive ability first suggested by Galton (1883) and Spearman (1904). However, the observation of this congruence should not be taken to imply that Irwin and Galton would have explained any such correlation in the same way. Irwin (1984) appeared to suggest that bright children felt generally more relaxed in testing situations and were able to master new tests more quickly, whereas Galton and Spearman thought that sensory discrimination ability formed the basis of individual differences in intelligence.

### **2.1.3 Nettelbeck, Edwards and Vreugdenhil (1986)**

Nettelbeck, Edwards and Vreugdenhil (1986) also attempted to replicate the AIT-cognitive ability study of Brand and Deary (1982). The discrimination involved in the task was essentially the same as that used by Brand and Deary, i.e. subjects were required to judge the order ('High-Low' or 'Low-High') of two tones played consecutively. The tones had the same frequencies and volume as those used by Brand and Deary (1982), but were played binaurally through headphones. Nettelbeck, Edwards and Vreugdenhil (1986) introduced a new mask for the AIT task; the white noise mask used by Brand and Deary (1982) and by Irwin (1984) was replaced by a mask consisting of alternating 15ms bursts of white noise and both target tones. Their new mask was used before, between and after the tones, for 1000ms in each case. Auditory inspection time thresholds were determined using the parameter estimation by sequential testing (PEST; Taylor and Creelman, 1967) technique, an adaptive staircase psychophysical procedure.

The 30 subjects tested by Nettelbeck, Edwards and Vreugdenhil (1986) were aged from 20 to 40 years, with an estimated mean IQ of 121 (SD 8, range 102-135). AIT data were presented for 29 of the 30 subjects, indicating that almost all subjects were deemed able to complete the AIT task discrimination to the satisfaction of the testers, though the authors stated that, "some Ss find this task to be inordinately difficult". The mean AIT was 117ms (SD 116, range 11 to 399 ms). Twelve cognitive ability tests were included in the study, and their mean correlation with auditory inspection time thresholds was -0.25. The cognitive ability tests included Raven's Advanced Progressive Matrices (RAPM), the Digit Span test from the Wechsler Adult Intelligence Scale-Revised (WAIS-R), and nine subtests from the Comprehensive Ability Battery (CAB; Hakstian and Cattell, 1982). Significant correlations were obtained between AIT and RAPM, Digit Span forward from the WAIS-R and Associative Memory, Inductive Reasoning and Verbal Ability from the CAB (mean correlation = -0.36, all  $p < 0.05$ , one-tailed, range -0.33 to -0.38). The correlation between AIT threshold estimates and thresholds on the standard two-lines visual inspection time task was 0.39 ( $p < 0.05$ , one-tailed).

Nettelbeck, Edwards and Vreugdenhil (1986) concluded that AIT did not have a particularly strong relationship to the primary abilities of *perceptual speed* or *speed of closure*. While recognising that their sample was too small to afford confident conclusions concerning the differences between the various IT-cognitive ability correlations, they interpreted their results to indicate that IT was associated with general intellectual capacity rather than with specific abilities. Theirs was the first study to indicate that, in a non-handicapped subject sample, there was a moderate amount of variance shared by visual and auditory forms of IT-type tasks.

#### **2.1.4 Problems with the original versions of the AIT Task**

Although the AIT task had the advantage of not being penetrable by strategies, such as

apparent movement detection, which have affected the visual inspection time task (Egan, 1986, 1991; Mackenzie and Bingham, 1985; Mackenzie and Cumming, 1986, Mackenzie, Molloy, Martin, Lovegrove and McNicol, 1991), the studies presented above raised some obvious difficulties with the original AIT task which did not occur with the standard visual IT task. AIT testing tended to result in a skewed distribution of scores, with most subjects obtaining very short AITs and a few subjects obtaining very long AITs. Typical of the results were those of Brand and Deary (1982), where the range of AITs was 6 to 160ms, but where the median was 10ms. One factor which caused this was the inclusion of mentally handicapped subjects (Brand and Deary, 1982). Second, very brief threshold values were possible because the relatively quiet white noise appeared to be an ineffective mask, and subjects reported being able to 'rehearse' the target tones even after the masking noise had begun. Third, some very long threshold estimates might have been obtained from subjects who were, essentially, 'tone deaf' to the pitch discrimination that was required in the task. Although subjects were sometimes pre-tested for pitch discrimination ability (Brand and Deary, 1982; Irwin, 1984), those subjects who found the basic discrimination in the AIT task very difficult were not always omitted from data analyses in the studies by Irwin (1984) and Nettelbeck, Edwards and Vreugdenhil (1986). This is perhaps analogous to allowing subjects with inadequate visual acuity to proceed with the visual IT task; it introduces pitch discrimination variance into a task which was designed to index speed of processing alone.

A further problem with the original AIT task was the fact that, at very short tone durations such as 5ms, there was less energy in the fundamentals of the stimulus tones and that energy was spread over a wider range of frequencies (Irwin, 1984). Therefore, as the duration of the tone pairs became shorter, the AIT task might have changed character, becoming sensitive to speed of auditory processing *and*, increasingly, to pitch discrimination.

Therefore, the AIT used by Brand and Deary and by Irwin (1984) task was unsatisfactory, and tended to result in low threshold estimates, probably because it allowed the subject some time to process more stimulus information in the gap between the two stimuli. Subjects were able to rehearse the first of the two tones in short term memory in the 500ms (approximately, in the study conducted by Brand and Deary, 1982) or 1000ms (Irwin, 1984) between the tones, and the task may have allowed rehearsal of the second tone because the mask was not effective. In the modified mask devised by Nettelbeck, Edwards and Vreugdenhil (1986), this factor remained a source of some concern, because these authors ensured that, even though their mask was a quickly-alternating 'warble' of the stimulus tones and white noise, target tones were always followed by 15ms of white noise.

## **2.2 Other studies of auditory information processing and intelligence**

Studies conducted by other researchers appear to be relevant to assessing whether speed of processing in the auditory modality has a significant association with cognitive ability. Raz, Willerman, Ingmundson and Hanlon (1983) and Raz and Willerman (1985) used auditory backward recognition masking tasks to estimate individual differences in speed of auditory processing in undergraduates and related these to individual differences in cognitive ability test scores. Saccuzzo, Larson and Rimland (1986) used the 'repetition test' devised by Tallal and Piercy (1973) to study the association between auditory processing speed and cognitive ability.

### **2.2.1 Auditory backward recognition masking and cognitive ability**

Raz, Willerman, Ingmundson and Hanlon (1983) used an auditory backward recognition masking task first devised for auditory psychophysical investigations by Massaro (1970, 1973; Massaro and Kahn, 1973). This task involved a single target tone, either 770 or 870 Hz, played for 20ms and with rise and fall times of 5ms. The key variable in the task was the interval between the tone and the backward mask, which was manipulated by the

experimenter. The backward masking sound was a tone which was intermediate in pitch between the two possible targets, i.e. 820 Hz. Subjects indicated whether the target was 'high' or 'low'. The number of correct identifications of target tones was taken as a measure of auditory processing speed. Massaro (1976) and Kallman and Massaro (1979) have indicated that the problem of some subjects not being able to achieve near-perfect asymptotic levels of performance existed on the above task, as it did in the AIT task of Brand and Deary (1982). As many as 42% of subjects, even after substantial amounts of practice, were unable to achieve better than chance responses.

In the first study in the report by Raz, Willerman, Ingmundson and Hanlon (1983), one subject was excluded because of hearing problems and two of the remaining 16 subjects were unable to achieve better than chance levels on the task discrimination. The 14 subjects who provided data for analysis comprised two groups of seven subjects who were separated widely on Scholastic Aptitude Test (SAT) scores. Twelve stimulus-mask intervals (called inter-stimulus intervals, or ISI, in the study), ranging from 0 to 1000ms, were used. The stimuli were recorded on tape and played through headphones at 80 dB. Despite the number of subjects and very skewed auditory thresholds, the difference in thresholds between the groups was significant, with superior auditory processing being found in the high ability group. The correlation between a logarithmic transformation of ISI thresholds and SAT class (high or low) was -0.53 ( $p$  approx 0.05). Crude estimates of musical experience were not significantly correlated with ISIs.

In a second experiment Raz, Willerman, Ingmundson and Hanlon (1983) tested high ( $n=9$ ) and average ( $n=11$ ) scorers on the Cattell Culture Fair Intelligence Test. Apart from some modifications of ISI range used and in the training of and feedback to subjects, the auditory processing test was similar to that used in the first study, except that a 30ms tone was used in addition to a 20ms target tone. With both tone conditions there was a clear ceiling effect for high ability subjects, i.e. many subjects did not require *any* ISI duration

in order to report accurately whether the first tone was 'high' or 'low'. Correlations between logarithmic transformations of ISIs estimated using the 20ms and 30ms tones and Cattell IQ scores were -0.69 and -0.73 (both  $p < 0.001$ ), respectively.

Raz, Willerman, Ingmundson and Hanlon (1983) discussed their results in terms of the 'noisier' information processing channels of lower IQ subjects, and remarked that their interpretation was congruent with the hypotheses of Hendrickson and Hendrickson (1980), who had suggested that the high IQ subject is characterised by a greater fidelity of neural transmission. They reckoned that the use of a 30ms in addition to a 20ms target tone in the task did not throw any more light on individual differences in auditory information processing because,

the 20ms tone was so easy for the high intelligence group that there was little room left for improvement of recognition speed with increased processing time.

A further problem with this task was that performance did not worsen steadily with lower ISIs; the relationship between ISI duration and percentage of correct responses in some subjects showed a U-shaped function. This occurrence had been recognised by Massaro (1970, 1972, 1976) and it was hypothesised that the superior performance before the "trough" was caused by a peripheral pitch shift of the target tone away from the masking tone that made target tone recognition easier at some short ISIs. This "dip" in performance did not occur in all subjects and occurred at different durations in different subjects. In the studies by Raz, Willerman, Ingmundson and Hanlon (1983), more low ability than high ability subjects had U-shaped functions, though the groups were too small to allow for statistical testing to achieve significant results.

A later study of the relationship between auditory information processing and cognitive ability by Raz and Willerman (1985) used a variation on the Massaro (1970) task. In the introduction to this study the authors stated that their previous research (Raz, Willerman,



Ingmundson and Hanlon, 1983) was a straightforward confirmation of the AIT-cognitive ability association first reported by Brand and Deary (1982). In the 1985 study, 36 psychology students with a mean Cattell IQ of 125.6 (SD 24) were recruited. The auditory information processing task introduced the following novel features when compared with their previous studies: stimulus tones were 870 and 770 Hz with a masking tone of 820 Hz; ISI varied from 0ms to 480ms; and one of three target tone durations, 10ms, 13ms and 20ms, was played to subjects before the masking tone on any one trial, instead of a single 20ms tone duration. Their intention, in having subjects make discriminations based upon target tones played at one of three randomly-selected durations, was to make the auditory task more loaded with requirements for selective attention and, thereby, to observe any changes auditory threshold or auditory threshold-IQ correlations that occurred as a result.

Correlations between the probit-derived thresholds and Cattell IQ scores were -0.37, -0.40 and -0.44 (all  $p < 0.05$ ) for the 10ms, 13ms and 20ms tone durations, respectively. Raz and Willerman (1985) concluded that the magnitude of the aptitude-related differences was not affected by the variation in target tone duration, though the number of subjects was too small to accept the null hypothesis with confidence. They also concluded that their stated attempt to 'add' a selective attention component to the task did not appear to have altered the task at all, and that the same processes were being tapped by this "more complex task" as had been tapped by their previous task (Raz, Willerman, Ingmundson and Hanlon, 1983).

The problem of subjects not being able to perform at above chance levels on auditory information processing tasks appeared again in the study by Raz and Willerman (1985). Six subjects could not score at better than chance levels on the 10ms target tone form of the task and some were responding at no better than chance levels on the 13ms and 20ms versions of the task. The authors' solution to this was to assign these individuals the longest stimulus onset asynchrony threshold obtained by any of the other subjects and to



add 1ms. This solution implies that the subjects responding at chance levels had very long stimulus time requirements, which might have been true, but it appears to ignore the more prosaic and, perhaps, more likely hypothesis that these subjects could not make pitch discriminations sufficiently well to be included in the data analyses. The problem of skewed distributions of auditory thresholds occurred in this study, and it is clear from their scattergrams that a few individuals had very aberrant thresholds.

### **2.2.2 The 'repetition test' and cognitive ability**

Saccuzzo, Larson and Rimland (1986) correlated visual and auditory processing speed measures with mental ability test scores in 96 students. The auditory tests used were forms of the 'repetition test' devised by Tallal and Piercy (1973). This test involved playing one of four possible combinations ('High-Low', 'Low-High', 'Low-Low' and 'High-High') of two tones of markedly different pitch (100 versus 305 Hz), and required subjects to indicate the pattern of tones that had been presented. Tones were complex and had a rise and fall times of less than a millisecond. Stimulus tone duration and interstimulus interval were varied in the original study by Tallal and Piercy (1973), where it was discovered that,

Total duration of stimulus patterns proved critical to aphasics' performance. It is suggested that developmental aphasics are incapable of perceiving auditory information at a normal rate, and the possibility is considered that this constraint on the speed of auditory processing may underlie their language impairment.

Therefore, the task devised by Tallal and Piercy (1973) has an empirical association with AIT in that it places emphasis on the stimulus time available for the early stages of auditory information processing. There is also a theoretical parallel with AIT, because the Tallal and Piercy theorised that the task was assessing individual differences in the *speed* of auditory processing and that these differences might underlie differences in language development. Tallal, Stark and Mellits (1985a) continued to research on the relation between the repetition test and the development of language in children and concluded that,

...auditory perceptual variables, specifically those requiring rapid temporal analysis, were most highly correlated with the degree of receptive language deficit of the dysphasic children

Further, Tallal, Stark and Mellits (1985b) extended their battery of rapid information processing tasks to the visual and tactile modalities; they asked subjects to discriminate the temporal order of light flashes and the location of simultaneous touches on both sides of the body. They found that a number of variables were predictive of poor language ability in children and that all of the variables that contributed significantly to a multiple regression equation predicting language problems,

assessed the ability to produce or perceive information either simultaneously or rapidly in succession, regardless of whether the information was verbal or nonverbal.

Saccuzzo, Larson and Rimland (1986) did not provide a full description of their version of the 'repetition test' auditory processing task. They indicated that subjects were asked to report the order of two-element tone patterns, that the dependent variables were the numbers of errors from the long- and short-ISI conditions of the test, and that these were taken to be indices of, "speed of auditory processing". Stimuli were played on a tape recorder. The mean of four correlations between the two auditory processing measures and two visual inspection time composite estimates was 0.21 ( $p < 0.05$ ), i.e. brighter subjects had faster auditory processing. A factor analysis of various information processing tasks revealed a general mental speed second order factor and separate first order factors labelled reaction time processing, auditory processing and visual processing. Correlations between the auditory processing task with the shorter ISI and SAT scores and Freshman Grade Point Averages were 0.23 and 0.25, respectively (both  $p < 0.05$ ), though there were non-significant correlations with High School Grade Point Averages, Block Design and vocabulary scores.

Saccuzzo, Larsen and Rimland (1986) found that IT (visual and auditory) and RT measures loaded on the same second order factor as did conventional paper-and-pencil ability test measures, but emphasised that IT tasks appeared to have general mental speed variance in addition to task-specific variance. They appeared sympathetic to the notion that a battery of information processing tasks might be used to index general mental ability, given that task-specific variance might be averaged out by such a procedure. They also concluded that they had preliminary evidence to indicate that visual IT tended to be more strongly associated with right hemisphere processing and that auditory processing speed tended to be related to left hemisphere ability.

## **2.3 Conclusions and a research plan**

### **2.3.1 Auditory information processing and cognitive ability test scores**

From the above review it appears that measures of auditory information processing estimated by psychophysical procedures tended to have significant but modestly-sized correlations with cognitive ability scores. Research reports using three versions of an auditory task that was designed to be analogous to the visual inspection time have resulted in correlations with cognitive ability scores that are in the same range as correlations between psychometric test scores and visual inspection time threshold estimates (Brand and Deary, 1982; Irwin, 1984; Nettelbeck, Edwards and Vreugdenhil, 1986). In what might be construed as confirmations of these reports, studies by Raz and colleagues (Raz, Willerman, Ingmundson and Hanlon, 1983; Raz and Willerman, 1985) obtained estimates of individual differences in the amount<sup>of</sup> silent time required after a target tone of either 770 Hz or 880 Hz in order to allow its accurate discrimination. These estimates correlated significantly with SAT scores and IQ-type measures. Saccuzzo, Larsen and Rimland (1986) also interpreted their results with the repetition test as confirmations of the

AIT-cognitive ability association.

The existing AIT studies are few in number and two of them involved small numbers of subjects. Moreover, it has not been established whether AIT correlates more closely with any particular type of mental ability. Therefore, as a first priority, studies in this thesis will attempt to discover whether the above associations remain when larger groups are tested and they will include tests of both verbal and non-verbal ability. Such attempts will be done in the context of other issues and, therefore, the studies in Chapters 4 to 8, inclusive, will be relevant to these issues.

### **2.3.2 Auditory and visual information processing**

The finding by Nettelbeck, Edwards and Vreugdenhil (1986), that there was a correlation of 0.39 between auditory and visual IT, is potentially important. It is a cross-task and cross-modal correlation that is relatively large by the standards of those other studies that have attempted to isolate 'components' of cognitive functioning and to establish that those putative components may be identified with sufficient reliability to allow them to be pinpointed as the source of shared variance that causes apparently different tasks to correlate significantly (see Chapter 1; Sternberg and Gardner, 1983). The hypothesis that visual inspection time might be indexing information apprehension speed that was general, i.e. cross-modal, did obtain some support from the study by Saccuzzo, Larsen and Rimland (1986) and, perhaps, from Tallal, Stark and Mellits (1985b). However, it did not obtain confirmation from the study by Irwin (1984), and the intercorrelation of the visual and auditory IT thresholds in Brand and Deary (1982) did not hold when two mentally handicapped subjects were excluded from the data analyses.

Therefore, in Chapter 4, auditory inspection time will be measured alongside recognised forms of the visual IT task in order to establish whether there is an association between speed of processing in the two modalities.



### 2.3.3 Modifying the auditory inspection time task

Some difficulties have arisen in the considerations of extant versions of the AIT task which appear to warrant an attempt to modify the task while retaining its speed of processing nature. A proportion of subjects appeared to find the discrimination of two tones separated by 110 Hz to be difficult. A similar finding has been reported by Massaro (1976; Kallman and Massaro, 1979) in a different auditory processing task that requires the same pitch discrimination. It has been suggested above that those subjects who find the attainment of near-perfect discrimination performance to be impossible even at long stimulus durations might be identified by their relatively poor pitch discrimination abilities on unspeeded pitch discrimination tasks. This hypothesis attained partial support in the study conducted by Irwin (1984), where AIT estimates were correlated significantly with Seashore pitch scores. Irwin's (1984) suggestion that pitch discrimination ability was a confounding factor in AIT task performance or, indeed, that pitch discrimination ability was the *only* ability indexed by AIT thresholds, was supported by his analyses of the very brief tone durations used in the AIT task; longer stimulus tone durations retained their separateness of pitch, whereas tones of 5ms duration appeared to be physically more similar in pitch.

Therefore, it was a priority in the present thesis to develop a form of the auditory inspection time task that was identifiably similar to that used in previous studies, but which did not result in most subjects having very brief threshold estimates. Such a new task will be described in Chapter 4, and it will be used in the studies reported in Chapters 4, 5, 7 and 8. It also remains important to discover whether any AIT-cognitive ability test score correlations fall to non-significant levels when individual differences in pitch discrimination are controlled for. This will be tested in various forms by the studies reported in Chapters 5, 6 and 7.

#### **2.3.4 Resurrecting a pitch discrimination-cognitive ability association?**

An intriguing and important hypothesis that follows from Irwin's (1984) study is the possibility that cognitive ability is related to individual differences in the ability to make simple pitch discriminations. This hypothesis was originally produced by Galton (1883) and, because it appeared to be clear to the present author that historians of psychology were unanimous in their agreement that subsequent empirical studies had proved Galton wrong, a re-examination and some re-analysis of the historical literature concerning cognitive ability and sensory discrimination was undertaken. This review and re-analysis will form the basis of Chapter 3. The hypothesis that pitch discrimination has a significant association with cognitive ability test scores will be tested in various forms in Chapters 5, 6 and 7.

#### **2.3.5 Development of AIT and its causal status with respect to cognitive ability**

Few studies have examined the development of visual inspection time (Nettelbeck and Wilson, 1985; Nettelbeck and Young, 1990; Anderson, 1988), and no studies have examined the development of auditory inspection time. In addition, longitudinal studies of visual inspection time and cognitive abilities across time have tended to be conducted with relatively few subjects and have not used formal structural modelling techniques to test hypotheses about the causal association between information processing and cognitive ability (Nettelbeck and Wilson, 1985; Nettelbeck and Young, 1990). This latter question is arguably the most important issue in IT research. Therefore, Chapter 8 will review the literature on longitudinal studies of visual IT and cognitive ability, and will present data from a moderately large scale longitudinal study of the changes in cognitive ability and AIT across two years from age 11 to age 13.



## Chapter Three

### Sensory discrimination and intelligence: historical review and re-analysis

*...no matter what the exact mechanisms of information processing underlying intelligence, Galton's (1883) suggestion of an important link between 'the avenue of the senses' and good sense may not be as far fetched as previously supposed (Raz, Willerman and Yama, 1987).*

#### 3.1 Introduction

A previous paper by this author detailed the historical precedents of inspection time (Deary, 1986), and it was concluded that there were sporadic findings, from the 1880s onward, to indicate that individual differences in speed of perception were associated with measures of intelligence. This finding appeared to have been unknown to most researchers in the field of intelligence research. Perceptual speed, though, was not a central concern of the early investigators who were interested in the links between intelligence and 'simpler' abilities. A more prominent aim of researchers in the late nineteenth and early twentieth centuries was to discover whether the ability to make fine discriminations in various sensory modalities was related to intellectual ability. The search for a consistent correlation between intelligence and measures of simple sensory discrimination appeared to have been fruitless, judging by the lack of interest present day differential psychologists have in just noticeable differences, and the negative evaluations of the original research efforts (see below).

In Chapter 2 it was shown that one of the explanations put forward to account for the relationship between AIT threshold differences and scores on tests of cognitive ability was that brighter subjects were better at making unspeeded simple discriminations involving a pitch difference between two tones. This appeared to be a surprising suggestion, since it



hypothesised that an association existed between mental ability as tested by psychometric tests and simple sensory discrimination abilities. This was similar to the sensory discrimination-general mental ability association originally proposed by Galton (1883), and generally held to be incorrect. These considerations prompted a re-examination of the history of attempts to associate sensory discrimination measures and mental ability test scores.

### **3.2 Current conceptions of the early studies**

Modern historians of intelligence rarely undertake their work without a bias or a practical task in mind. Psychologists writing an introductory text which includes a chapter on intelligence often provide a quick summary of the history of intelligence research for their readers. Thus, Bernstein, Srull, Roy and Wickens (1988), for example, stated that,

In the late 1800s, Sir Francis Galton tried, unsuccessfully, to develop a test of intellectual ability by measuring people's perceptual and motor abilities, such as how fast they responded to simple stimuli and how sensitive they were to pain. Other researchers soon concluded that these abilities had very little to do with intelligent behaviour (Wissler, 1901).

Others have selected certain episodes and findings in the history of intelligence research to develop a particular thesis. For instance, in developing his arguments against the hereditarian position with regard to mental ability, Gould (1981) remarked of the first efforts to test for individual differences in mental abilities that,

Galton, without notable success, had experimented with a series of measurements, mostly records of physiology and reaction time, rather than tests of reasoning. Binet decided to construct a set of tasks that might assess various aspects of reasoning more directly.

As will be demonstrated, from the number of early studies, the length of many of them and the diversity of theorising therein it is possible to present the research evidence in many different ways. Nevertheless, there is some broad agreement in the following areas, viz. the studies that were important, the results that were obtained by these studies, and the

theoretical orientations of the early workers.

Blum (1978), commenting on Galton's efforts with thousands of subjects in the 'Anthropometric Laboratory', reckoned that,

He [Galton] could not find any clear relationship between simple sensory acuities and the more global phenomenon of mental ability... J. McKeen Cattell, who studied with Galton, attempted similar kinds of measurements in the United States, and elaborate statistical analysis of his data failed to show any dependable relationships with course grades of college freshmen... Galton's only major conclusion from his own studies was that women on the average were inferior to men in every respect... Binet's tests correlated with performance in school, whereas Galton's and Cattell's did not.

In this short account many of the major players and themes have appeared, as have some inaccuracies. Spearman, in the review that begins his 1904 paper, wrote of Galton's attempts to correlate the "psychical tendencies" with mental tests,

[Galton was] a suggestive writer [who] appears to have been diverted from the point by other interests, and to have contented himself with the above general impression without clinching the matter in systematic investigation.

The general impression to which Spearman (1904) refers was Galton's notion that,

...men of marked ability [appear] to possess on the whole an unusually fine discrimination of minute differences in weight.

Carroll (1982) noted, contrary to the impression given in Blum's (1978) account, that Galton made very little use of correlation - his own invention - in his anthropometric work. In fact, a scrutiny<sup>of</sup> Galton's *Inquiries into Human Faculty* (1883) revealed no application<sup>of</sup> statistical analysis to his work on mental ability; his conclusions were made using impressions and anecdotes,

The discriminative faculty of idiots is curiously low; they hardly distinguish between heat and cold and their sense of pain is so obtuse that some of the more idiotic seem hardly to know what it is. In their dull lives such pain as can be excited in them may literally be accepted with a welcome surprise... I saw a boy

with the scar of a severe wound on his wrist; the story being that he had first burned himself slightly by accident, and, liking the keenness of the new sensation, he took the next opportunity of repeating the experience, but, idiot-like, he overdid it.

Similarly, when it came to making his statement on sex differences in sensory acuities and their relation to general mental ability Galton (1883) offered no statistical analysis,

The trials I have as yet made on the sensitivity of different persons confirms the reasonable expectation that it would on the whole be highest among the intellectually ablest. At first, owing to my confusing the quality of which I am speaking with that of nervous irritability, I fancied that women of delicate nerves who are distressed by noise, sunshine, etc., would have acute powers of discrimination. But I found this not to be the case. In morbidly sensitive persons both pain and sensation are induced by lower stimuli than in the healthy, but the number of just perceptible grades of sensation between them is not necessarily different. ...as a rule... men have more delicate powers of discrimination than women, and the business experience of life seems to confirm this view. The tuners of pianofortes are men, and so I understand are the tasters of tea and wine, the sorters of wool, and the like... Ladies rarely distinguish the merits of wine at the dinner table, and though custom allows them to preside at the breakfast-table, men think them on the whole to be far from successful makers of tea and coffee.

Because Galton's observations were so anecdotal and unscientific it is easy to misrepresent them. However bad Galton's evidence was for his conclusions, the quotation above clearly focusses on individual differences in just noticeable differences in sensations as the basis of the intellect. Therefore, Sternberg (1990) was inaccurate in his interpretation of Galton's discrimination-based theory of intelligence when he stated that,

He [Galton] also discovered that people are inferior to cats in their ability to perceive tones of high pitch. This finding presents a problem for any psychophysically based theory of intelligence that subscribes to a notion of evolutionary continuity. This suggests that, in at least this one respect, cats are superior in intelligence to humans.

If Galton may not be looked to for the empirical test of the hypothesis that there is a positive correlation between sensory discrimination and intelligence, then who else is cited as having failed in the search for such an association? Writers referring to the failure of this endeavour generally refer to two, often deemed influential, studies: those of Clark

Wissler (1901) supervised by McKeen Cattell at Columbia, and of Stella Sharp (1898-9) overseen by Titchener at Cornell. With regard to their effect on the hypothesis that there was a relationship between simple sensory measures and mental ability, Fancher (1985b) wrote of, "Wissler's devastating results," and Eckberg (1979) stated that Wissler's results were, "so dismal that they directly caused Cattell to end his own involvement with testing." Similarly, Carroll (1982), after mentioning the studies carried out by Wissler on Cattell's data at Columbia and by Stella Sharp, commented,

From the debates in the literature one would think that the mental testing movement was being laid to rest.

Carroll (1982) underlined the dead-end nature of the sensory discrimination approach to individual differences in intelligence with his comment on the applicability of the measures used at that time,

In the early years ...use was made of relatively simple tasks, usually involving powers of sensory acuity and judgement (e.g. detecting small differences in the weights of two visually similar objects) or speed of reaction time in responding to stimuli (e.g. naming colours). ...perhaps only one has survived in the current measures of intelligence - the memory span test.

Further evidence for the centrality of the Sharp (1898-9) and Wissler (1901) studies came from Herrnstein and Boring's (1965) *Source Book in the History of Psychology*. This book dealt with central issues in the history of psychology by gathering extracts from some of the most important papers in the area. In the section entitled 'Evolution and Individual Differences,' apart from sections by Darwin and Ebbinghaus, they included portions from *Hereditary Genius* by Francis Galton, and the Wissler (1901), Sharp (1898-9) and Spearman (1904) papers. These were the only papers included in the section and they were presented as the most influential contributions to the debate at the time. Writers who discuss the Sharp and Wissler studies often do not appear to doubt that the generally negative results they produced were responsible for a decline in interest of the search for a simple sensory measure of intellectual functioning.

Given the established importance of these studies, it is interesting to note how they are presented by various recent writers, and to reassess the studies themselves. The commentators separate roughly into two groups; those who accept that the results demonstrated the uselessness of the search for simple sensory correlates of intelligence, and those who are more critical of these putative key studies.

The first group to be examined is the broadly uncritical group. In this category there is the account of the simple sensory work that was offered by Forrest (1974) in his biography of Sir Francis Galton,

The supposition that sensory acuity bears some relationship to intellectual ability was destroyed by the work of Wissler and others at the turn of the century. This research also showed that sensory measures do not themselves intercorrelate, so that one cannot argue from the acuity of one sense to that of another.

Forrest (1974) gave no indication who the other workers might be and the reader has little alternative but to conclude that the matter was settled. Boring (1950) may be entered into the uncritical group with his account of the importance of these two studies. After discussing the Galtonian approach of studying the simple sensory underpinnings of intelligence he went on to look at the putative alternative as put forward by Binet.

On the other hand, Binet and Henri, as we have seen, invented tests of what are sometimes called higher mental functions... These tests touched more closely the faculties that everyone wished to assess, the abilities that make for success in life. Since educational psychology was supplying some of the motivation for the development of testing, success in school became for the time being the most desirable outcome to predict. As early as 1898 Stella Sharp at Cornell was able to show that Binet had won out over Cattell - if we may put this complex matter so simply. Sharp's conclusion was a decision of Titchener's laboratory that the Wundtian variables of experimental psychology are less adequate for a description of those human abilities that make for success than are Binet's devices, which he made up and did not, in general, come directly out of the laboratories. Perhaps Titchener felt even then - as he did later - that applied psychology is scientifically unworthy and that failure of 'pure' experimental psychology to meet the requirements of functional use was not disparaging to the Wundtian school. Or perhaps he was glad to find Cattell in error.



Boring (1950) did not raise the possibility that Sharp's study was not sound enough to draw such broad conclusions from. Sharp's (1898-9) introduction was more careful than her methodology. She discussed whether the psychologist interested in individual differences should use the same processes as the experimental psychologist to, "account for unlike results from the building up of unlike materials." Her answer was that the French reckoned yes, the Germans no, and that the American psychologists did not know. She stated that her study was an attempt to implement the tests suggested by Binet, and her investigation involved *no* simple sensory measures. In summary, she concluded that she had found no worthwhile results and that the methods of higher mental function testing and simple sensory testing should be combined. Therefore, Boring's (1950) evaluation of Sharp's (1898-9) study is puzzling, and would be unlikely to have been drawn by anyone who had scrutinised Sharp's tests, subject population and her lack of statistical analysis or conclusions; all of these considerations make it impossible to support Boring's summary. Sharp (1898-9) did, however, express doubts about the entire mental testing enterprise,

So many part-processes are involved in the complex activities and the manner of their variation so indefinite, that it is seldom possible to tell with certainty what part of the total result is due to any particular component. It is doubtful if even the most rigorous and exhaustive analysis of test results would yield information of importance as regards the structure of mind.

Boring (1950) appended no critical comments to his summary of Clark Wissler's results of 1901,

...the correlation of class standing with reaction time was only -0.02 and with the test of logical memory only +0.16, whereas the correlation of class standing with gymnasium score was +0.53, and of performance in Latin with performance in Greek +0.75.

With Boring's (1950) raising of the possibility that Titchener's own preferences might have affected the way in which he greeted the results, we have an indication that this is a topic where the discourse often slips outside the expected mode of disinterested scientific comment. The following is Eckberg's (1979) obituary of the efforts to find simple

correlates of mental ability,

A final study, which proved a staggering blow to the testing movement, was performed by Clark Wissler at the Columbia University Laboratories, then under the direction of Cattell. Wissler employed 21 different tests, an elaboration of the tests suggested by Cattell a decade previously. He then used Pearson's new method of correlation to determine the relationships among psychological tests, anthropometric measures, and college grades. The results were so dismal that they caused Cattell to end his own involvement with testing.

It was with the Wissler study that testing entered its period of decline. R.D. Tuddenham (1962) reports that by 1905 the academic movement was "moribund," so much so that Binet's later work almost did not revive it. The movement waned as a result of a decade of failure to discover important hierarchical mental differences. The negative findings of Sharp and Wissler are commonly cited as the crushing blows to the movement (for example Peterson 1925, Tuddenham 1962), but their studies can better be seen as the culmination of a long line of studies that hardly ever produced the kinds of results expected. Be that as it may, the enthusiasm of testers declined markedly after 1901.

The metaphorical language is rather odd in the context of scientific evaluation, i.e.

"staggering blow," and "crushing blows". However, there are also errors in the above.

First, no one, until the work of Spearman in 1904, was in a position to look for a hierarchy of mental abilities because the concept had not been articulated and the statistical sophistication barely existed. Second, it is not clear to whom Eckberg (1979) was referring when he stated that Sharp and Wissler were the last in a long line of studies that failed to deliver the expected findings. Until the Wissler (1901) study no-one had performed a correlation between any two tests that were relevant to the search for the simple measures underlying mental ability. Sharp's results were entirely impressionistic, her tiny sample was very restricted in intellectual range and not a single correlation was computed. Eckberg (1979) stated that,

In one widely cited study, Stella Sharp (1899) found little consistency among schoolchildren as they moved from test to test.

In addition, Sternberg (1990) remarked that,

Sharp (1899) undertook a large-scale experiment to discover the usefulness of the Binet-Simon tests in applied settings.



In fact, Sharp tested only *seven postgraduate students*. Moreover, neither the simple sensory/intelligence movement nor the Binet-type approach were anything like moribund at the time. With respect to the former, it cannot be unusual for a single laboratory to give up a line of enquiry after a long study, inadequately designed, has not yielded helpful results. However, in the UK the work of Spearman (1904) at the University of London and the work of Burt (1909) were begun just after this time. Concerning the rise of the Binet approach, there was no need to perform the risky resurrection that Eckberg (1979) tried to imply took place. As Boring (1950) documented,

...Thorndike was in a position in his *Educational Psychology* of 1903 to show what kinds of tests were best for predicting educational success. At the end of this decade Goddard had gotten out his own revision of the Binet-Simon scale, and Whipple had published the first edition of his *Manual of Mental and Physical Tests* with his description of fifty- four tests and how to give them. By 1910 mental testing had clearly come to stay.

Eckberg's patient appears to have rallied after the diagnosis of a terminal illness.

Although Fancher (1985b) belongs in the group of researchers who have been largely uncritical of the Wissler (1901) and Sharp (1898-9) studies, in his account of the Wissler study he does give a helpful account of the tests used by Wissler. The summary of the results has a funereal tone and Fancher's (1985b) language uses the same set of physical injury analogies as Eckberg (1979).

During the decade of the 1890s the cause of mental testing was enthusiastically taken up by an increasing number of investigators in several different countries. Gradually, however, it became evident that there was something seriously wrong with the tests which did not really seem to measure useful differences in "mental" functions, as they had been designed to do. The crowning blow was struck in 1901 by Clark Wissler, one of Cattell's own graduate students, who obtained both mental test scores and academic grades from more than 300 Columbia University and Barnard College students. Wissler also learned the techniques for computing correlation coefficients, just recently perfected by Karl Pearson, and so was able to estimate with mathematical precision the exact interrelationships between the various mental tests, and independent measures of intellectual achievement.

Wissler's devastating results indicated that the "mental tests" showed virtually *no*

tendency to correlate with academic achievement; for example, class standing correlated  $-.02$  with reaction time,  $+.02$  with colour naming,  $-.08$  with dynamometer strength, and  $+.16$  with memory for number lists.

Wissler's results greatly disappointed psychologists. Perhaps realising what his research had done to psychology, Wissler shortly switched fields to become an anthropologist and one of the earliest American supporters of the environmentalist "culture concept" explanation for differences between ethnic groups. Cattell remained a psychologist but lost much of his enthusiasm for the Galtonian approach to mental testing, and gradually turned his primary attention to scientific administration and the editing of journals.

Thus Fancher (1985b) portrays Wissler as having set up a model study with hundreds of subjects, new statistical tests providing mathematical precision and a series of independent measures with which to correlate the simple sensory and other measures.

Let us now turn to the accounts that are more critical of the Sharp (1898-9) and Wissler (1901) studies. Jensen, in his *Bias in Mental Testing* (1980) gave a subtly different account to that of Fancher's,

A historically fateful study by Clark Wissler (1870-1949), one of Cattell's own Ph.D. students in Psychology at Columbia University, effectively signalled the demise of the Cattell battery of tests as a measure of intelligence.

Wissler became the first psychologist actually to compute a Pearsonian coefficient of correlation between RT and "intelligence." What he found was a disappointment: a correlation of only  $-0.02$  based on 227 male students in Columbia College. The meagre results of an apparently carefully executed study, which was conducted in the country's then most prestigious psychological laboratory, was so singularly unimpressive and became so widely cited as to throw a pall over the investigation of RT as a potential means for measuring intelligence.

Jensen (1980), while conceding that Wissler's (1901) study might have been the most methodologically sound study of its time and the most statistically sophisticated, did indicate that a close look at the study made it virtually useless and, in fact, incapable of deciding the issue it set out to investigate. First, Wissler calculated a small fraction of the correlations that were possible, 42 out of over 600. Of course, correlations were burdensome and prone to error in the days before calculators and computers, as Fancher (1985a) has demonstrated in a careful re-analysis of Spearman's work of 1904. Second,

although there was no realisation of it at the time, Wissler's subjects were restricted in ability range, making it unlikely that he would obtain a decisive result. Not only were the subjects university students, they were students at an Ivy League university, and the restriction of range would have lowered any correlation considerably. Third, no attempt was made to examine the correlation matrix to look for regularities in the results. Finally, no account was taken of the errors in the individual measurements. Wissler's (1901) study had used only simple RT to sound (whereas his mentor, McKeen Cattell (1890), had written a decade earlier that RT, "was essentially a reflex... some measure more purely mental should be measured.") and, worse, had used only 3-5 reactions. Jensen (1985) reported that on Berkeley students the test-retest reliability of reaction times based on so few trials was about 0.35.

In their account of the Wissler and Sharp studies, Cairns and Ornstein (1979) accepted the Wissler (1901) results, but they were more critical of Sharp's (1898-9) study, which is often cited to demonstrate the superiority of the Binetian over the Galtonian approach to intelligence research,

Sharp (1899), a student working under Titchener's direction at Cornell, was also pessimistic about the "method of tests." Sharp concluded that Binet's tests of complex functions were more effective than tests of simple psychological function as devices for differentiating among individuals.

But she then went on to argue that there was relatively little correspondence among these various tests and that much additional work would be necessary before this approach would be of real value in practical situations. Further, Sharp stated that testing would not contribute much to structural psychology. Despite the fact that Sharp's study was basically inadequate to the task of evaluating the Binet approach - she tested only seven subjects (all advanced students in psychology), with materials and under conditions different from those recommended by Binet - her report was influential and had the effect of dampening interest in the assessment of intelligence.

Of course, it should now be obvious that neither the Sharp (1898-9) nor the Wissler (1901) study were adequately designed for testing the Binet or the Cattell/Galton approaches to intelligence, as they have been dubbed. To be sure, the Wissler (1901)

study did produce negative results but, as far as they produced any results at all, the tests of Sharp (1898-9) were actually a failure to implement the Binet methodology, not a vindication as some writers have suggested.

Watson's (1979) summary of the research that has been considered so far provides a suitable conclusion to this section,

Most of the investigations of the time were concerned with simple sensorimotor and associative functions and were based on the assumption that intelligence could be reduced to sensations and motor speed, an attempt which, as is now known [sic], was doomed to failure. Furthermore, although more suitable verbal material was used, the studies of College students at Cornell, such as Sharp's, and the Wissler study at Columbia, were found to be essentially nonpredictive. What the workers failed to take into account was the fact that college students are a highly selected group having a considerably restricted range. The negative findings of these studies effectively blocked further investigation at the college level for years. When one stops to consider that the dominant systematic position of the day was the structuralism of Titchener, who had banished tests as nonscientific, it is no wonder that tests were viewed with at least a touch of condescension.

It was on this very weak empirical basis that the two influential American University studies are reckoned to have dissuaded further research on the hypothesis that simple sensory measures form the bases of cognitive ability. In fact, more decisive studies in the search for the simple sensory correlates of intelligence were carried out by two British psychologists and were published in prestigious journals. By the time these studies were reported the Binet test had been constructed and its immediate practical applications probably diverted the attention of psychologists away from the simple sensory correlates of intelligence. Nevertheless, this does not mean that the results of Spearman (1904) or Burt (1909-10) were any the less important. The fact that their instruments were less convenient than those of Binet is irrelevant to deciding whether, in the end, the Galton/Cattell approach was vindicated.

### **3.3 The studies of Spearman and Burt**

It was not possible to resolve the question of the relationship between simple sensory

measures and general intelligence using the Wissler (1901) and Sharp (1898-9) data. In fact, in Spearman's well-known 1904 "General Intelligence.." paper he opened with a near-exhaustive literature review of attempts to locate the elements of intelligence in the senses, and he concluded that he had found little of substance and much poor experimentation and contradictory evidence. There was very little quantitative statistical examination of data in the studies Spearman (1904) reviewed. Most studies before Wissler's (1901) had used qualitative impressions of the data (including Sharp (1898-9)) to inform their conclusions. Spearman's (1904) stated intention was to launch a "Correlational Psychology" the purpose of which was in,

positively determining all psychical tendencies and in particular those which connect together the so-called 'mental tests' with psychical activities of greater generality and interest.

Spearman's (1904) concerns even then seemed to anticipate the fact that a focus on testing that was not tied to basic psychological processes would alienate the study of intelligence from experimental psychology,

Binet and Henri appear now to seek tests of a more intermediate character, sacrificing much of the elementariness, but gaining greatly in approximation to the events of ordinary life. The result would seem likely to have more practical than theoretical value.

Spearman's (1904) review merits some consideration as it placed the work of Wissler and Sharp in the context of other attempts to settle the issue of whether more simple sensory and motor abilities were related to higher cognitive functions. Spearman (1904) cited Binet and Vaschilde (1897) who had tested 455 schoolchildren, average age 12, on reaction time, choice reaction time, memory for numbers and the number of dots that could be made on a piece of paper in 5 seconds. "The intellectual order," reported Spearman (1904), "harmonises badly with reaction times and harmonises well with the memory for numbers." Better than either, though, was the ability to make dots. An attempted replication of the result, using 43 student teachers, yielded a relationship between memory



and dot-making that was inverse. This latter sample, though, would have been of a more restricted in range of ability.

Spearman (1904) discussed a study conducted by Seashore (1897-99) who rated teachers' estimates of intelligence with memory for time, pitch discrimination, loudness discrimination and illusions of form, colour and weight in 200 children, and who had found "little functional relation". Bagley's (1901) study (Spearman, 1904) of 160 schoolchildren found that reaction time was not related to school intelligence, and that dotting ability might be inversely related. Bagley concluded that there was some "antagonism" between physical and mental tests. A study of 1507 children by Carman (1898-9; Spearman, 1904) examined the relationship between pain sensibility and hand strength, and whether children were estimated to be bright or dull. Bright boys were more sensitive than dull boys, and those girls who were sensitive and stronger tended to be brighter. Carman, though, was not content with such general impressions and Spearman (1904) provided the reader with some of the more detailed results,

Those reported as being especially dull in mathematics were more sensitive on the right temple than on the left. Girls with light hair and blue or gray eyes are less sensitive to pain on the left temple but on the right temple they are more sensitive than the dark.

As Spearman (1904) commented, "This information is very curious."

Spearman (1904) discussed a further Binet study which tested 11 subjects who were the 5 cleverest and "six most stupid of a class of 32," and found no relation between reaction time, with or without stimulus choice, and intelligence. In this study intelligence did not seem to be related to dotting ability or the ability to spot changes in a metronome beat. However, intelligence did seem to be related to memory for numbers, and more intelligent children were able to extract more from a single glance at a piece of text. But, "the fullest correspondence of all," reported Spearman, " was presented by the very old test of Tactile



Discrimination."

Spearman (1904) judged that Sharp's (1898-9) study was not very encouraging, but neither was it considered by Spearman to be adequate, with such small subject sample and restricted ability range. Of Wissler's (1901) study Spearman (1904) remarked, "The final conclusions are about as blankly negative as could well be imagined." He further remarked that,

There is scarcely one positive conclusion concerning the correlation between mental tests and independent practical estimates that has not with equal force been flatly contradicted...

...in spite of the many previous inconclusive and negatory verdicts, the question of correspondence between the tests of the laboratory and the intelligence and will cannot yet be regarded as definitely closed. The only thing so far demonstrated is that the old means of investigation are entirely inadequate.

In summarising his literature review Spearman (1904) made four specific points about the pre-1904 investigations: only Wissler's (1901) study had used correlation coefficients, the others had relied on the impressions from the tabulated results; no investigators had given probable errors; "...in no case has there been any clear explicit definition of the problem to be resolved," since most studies had lumped in schoolchildren of a variety of ages and tried to, "kill as many birds as possible with one stone"; and no investigator took into account errors of observation.

Spearman's (1904) study appears to be the first attempt to test the relation between intelligence and simple sensory measures that approaches the methodological adequacy required to assess the outcome of the endeavour with any validity. The previous studies, though, were not completely useless. Spearman (1904) used these to inform his decisions as to which tests might be the most suitable. In one paragraph he set out his approach,

As regards the nature of the selected Laboratory Psychics, the guiding principle has been the opposite to that of Binet and Ebbinghaus. The practical advantages proffered by their more complex mental operations have been unreservedly

rejected in favour of the theoretical gain promised by the utmost simplicity and unequivocality; there has been no search after condensed psychological extracts to be on occasion conveniently substituted for regular examinations; regardless of all useful application, that form of physical activity has been chosen which introspectively appeared to me as the simplest and yet pre-eminently intellective. This is the act of distinguishing one sensation from another.

Thus Spearman's own account (1904) denied that there was any 'contest' between the Binet approach and the simple sensory endeavour. The two approaches attempted to do, and still do, different things. On the one hand there was and is the practical utility of a standardised test that might predict a person's ability to perform a number of mental operations and life tasks. On the other hand there was understandable interest in the basic processes, if there were any, that contributed to variance in higher mental abilities.

Spearman (1904) used tests of sensory discrimination in three modalities: auditory (pitch discrimination), visual (hue discrimination) and tactile (weight discrimination). For testing pitch discrimination Spearman used a "monochord" that was accurate to "one third of a vibration." Two notes were struck, one following  $\frac{3}{4}$  of a second after the other, and the subject was asked which was the higher tone. In the visual discrimination test the subject was asked to say which of two grey cards was darker. For weight discrimination Spearman used Galton's cartridges: two identical-looking cartridges were placed before the subject who picked them up and decided which was heavier. Spearman rejected the psychophysical methods of 'minimal changes' and 'right and wrong cases' in favour of a threshold measure. In all three tasks subjects were given some "fore-exercise" and an estimate of their threshold was made (from Spearman's description this was, essentially, the point at which 80% of discriminations were correct); "there is a steady progression from greater to smaller intervals until eventually a threshold is found where he can just give eight answers out of ten." Spearman then checked that more errors (>20%) were made at more difficult discriminations and that at easier discriminations the subject was scoring nearer to 100%.

Estimates of intelligence were rough-and-ready at that time. There were no IQ-type tests as such, and Spearman (1904) had to decide who should rate intelligence and whether there were different types of intelligence. School work, he thought, was "present efficiency" which, if corrected for age, would leave a measure of "native capacity." Peer ratings by the most able of the children were said to reveal "common sense" and the general impression that children made upon others allowed them to be classified as "bright," "average" or "dull." Spearman's (1904) investigation studied five samples: the 24 oldest children from a local school in Berkshire; the next 36 oldest from the same school; 24 boys from a preparatory school for Harrow; the same group again; and 26 male and female adults. The preparatory school boys were tested at short notice and in a group fashion, whereas subjects belonging to other groups were tested individually. For the village school subject sample there were nine correlations given between three estimates of intelligence and the three discrimination measures. Using Fancher's (1985a) recalculations (owing to minor errors in Spearman's original hand-calculated correlations) the mean correlation between discrimination and intelligence estimates was 0.39 (range 0.25 to 0.47), i.e. subjects estimated to be more intelligent were able to make finer discriminations. The correlations among the different tests of sensory discrimination were as follows: pitch versus light, -0.02; pitch versus weight, 0.41; light versus weight, 0.30. Spearman estimated that the correlation between general discrimination and general intelligence approached *unity* when corrections for unreliability of the tests were made. The above-presented correlations were not corrected for unreliability. In the preparatory school sample pitch discrimination was correlated with Classics, French, English and Mathematics marks to give an average correlation of 0.38 (range 0.27 to 0.44). The adult sample included no measure of intelligence.

With the contradictory results obtained from unsatisfactory studies behind him, Spearman's (1904) discussion was celebratory. Some of his remarks might easily be deemed incautious and, with regard to the purported identity between general

discrimination and general intelligence, Spearman cooled his claim in a later footnote that appeared in the Burt (1909-10) study,

This conclusion of mine was badly worded. I did not mean (as others have naturally taken it) that general intelligence was based on sensory discrimination; if anything *vice versa*. I take both the sensory discrimination and the manifestations leading a teacher to impute general intelligence to be based on some deeper fundamental cause...

Other claims made by Spearman (1904) in his discussion were not retracted and led to the doctrine of *g* which still remains today,

[The general factor] if it be mental at all, it must inevitably be one of the fundamental pillars of any psychological system claiming to accord with actual fact - and the majority of prevalent theories may have difficulty in reckoning with it.

...thus we are becoming able to give a precise arithmetical limitation to the famous assertion that "at bottom, the Great Man is ever the same kind of thing." This Central Function, whatever it may be, is hardly anywhere more prominent than in the simple act of discriminating two nearly identical tones.

Thus did pitch discrimination achieve its zenith as an explanatory variable for individual differences in general mental ability. Spearman took the statistical analysis no further in the 1904 study. There was no factor analytic method available and his claim to have discovered *g* was based upon the universally positive nature of the correlations and the idea of a hierarchy of correlations. In fact, Fancher (1985a) has shown that, when Spearman's arithmetical errors were corrected, the correlation hierarchy is less perfect than was recorded in the 1904 paper. Nevertheless, here was the first study of imputed intelligence versus simple sensory measures which might be said to have approached minimal standards of methodological and statistical adequacy, and it obtained modest and consistent correlations in the expected direction. In the latter part of his discussion Spearman (1904) re-examined the studies that most resembled his own and concluded that, while Gilbert's (in Spearman, 1904) data showed little correlation between discrimination and intelligence, Seashore's data (recalculated by Spearman) revealed a correlation of 0.24 between discrimination and intelligence. Spearman undertook a final reconsideration of



Wissler's (1901) study and decided that the sample was too homogeneous, the discrimination tasks were too like memory tasks and that the subjects were often tested unsatisfactorily. They were tested three at a time, by students in the department. As a final comment on the Wissler (1901) study Spearman (1904) offered the following observations,

...no less than 22 tests were carried out, many of a most difficult character, besides measuring the length and breadth of each reagent's head; that during the leisure moments afforded him in the course of these tests the observing "student officer of the department" had to note in writing the contour of the reagent's forehead, the character of his hair, the nature of his complexion, the colour of his eyes, the shape of his nose, the description of his ears, of his lips, of his hands, of his fingers, of his face, and of his head - and that this whole procedure is considered to be satisfactorily completed in forty-five minutes.

Burt published his first major empirical work in 1909-10. This was an extension of Spearman's study and Burt acknowledged help from Spearman and, from various footnotes in the article, it is clear that Spearman had seen the typescript before publication. There is no reason to discount this empirical work of Burt, as it is prudent to do with his later work (Hearnshaw, 1979). During the gathering of data J.C. Flugel tested the subjects in parallel with Burt, and the correlations for the two researchers' data were presented separately. In general, the Flugel correlations were at least as high as those of Burt, and often higher (in the same direction). Burt (1909-10) gave two reasons for carrying out his study. He reckoned that general intelligence, which was "above all supreme" in its importance, was under-researched; "the notice it has received from psychologists has been in proportion astonishingly scant." He rehearsed the same complaints of earlier studies as did Spearman (1904), but his intention appears to have been more like that of Binet's. Burt (1909-10) attempted to,

...determine whether higher mental functions would not show a yet closer correlation with 'General Intelligence' than was shown by simpler mental functions, such as sensory discrimination and motor reaction, with which previous investigations have been so largely engrossed.

Nevertheless, Burt (1909-10) included simple sensory measures among his tests. He did not emphasise Spearman's (1904) results, and when he cited those who thought that intelligence had a basis in sensory discrimination he chose Titchener as his exemplar. Burt's subjects were drawn from a "high-class" preparatory school in Oxford (the Dragon School) and from a local school whose children were the sons of local businessmen.

The experimental tests were presented in five categories. Sensory tests measured: two-point skin discrimination; the discrimination of two lifted weights; pitch discrimination; and comparisons of the lengths of lines by eye. Motor tests measured tapping speed and card dealing. Sensory-motor tests examined card sorting ability (by colour) and alphabet sorting (constructing a single ordered alphabet from two jumbled ones). Association tests measured: the immediate retention of concrete words, abstract words and nonsense syllables; the ability to do mirror drawing; and the ability to reproduce a pattern of dots after a series of brief tachistoscopic presentations (the Spot Pattern Test referred to in Chapter 1). One single test of "voluntary attention" measured the ability to make a pencil mark in the centre of circles printed irregularly on a moving tape. As in Spearman's (1904) study, the measures of intelligence would be considered unsatisfactory today, being taken from the impressions of the masters at the schools. Any doubt that the headmaster had concerning the rank ordering of two boys was to be resolved by the teacher asking himself the question, "Which boy is quickest at seeing the point of anything?"

In Burt's (1909-10) study, two-point discrimination and the discrimination of weights correlated at near-to-zero levels with intelligence estimates. However, pitch discrimination (using a method very like that of Spearman (1904)) correlated at 0.40 and 0.37 with intelligence in the two groups. (When corrected, these gave correlations of 0.52 and 0.41). The ability to discriminate the lengths of two lines correlated on average at 0.29 (using the average crude error) and at 0.31 (using the mean variable error) with estimates



of intelligence. As Spearman (1904) had done before him, Burt found that, of the sensory discrimination measures, it was pitch discrimination that correlated best with intelligence. Having made little mention of Spearman's (1904) claims regarding sensory discrimination in his introduction, and having announced his greater interest in higher mental functions, Burt (1909-10) appeared surprised by these results,

...before actually calculating the coefficients [we] believed we were finding no correlation throughout the sensory region. General Intelligence, then, shows little or no relation to senses which to civilised man are of low cognitive value; but it shows a marked relation to those senses which aid the perception of relations or formation of concepts, and are of high cognitive value.

Dotting ability correlated at 0.47 and 0.41 with intelligence estimates in the two groups; card dealing at 0.49 and 0.29; card sorting at 0.52 and 0.56; and alphabet sorting at 0.50 and 0.61. Burt (1909-10), at this intermediate stage in his climb up to the higher functions concluded that,

...tests combining perception with motor reaction seem to involve intelligence to a still higher degree than relatively simple sensory or motor tests.

Among the tests of the "higher mental functions", immediate memory for words correlated at 0.5 with intelligence estimates and at 0.67 with exam standing; mirror drawing correlated at 0.67 and at 0.54 with headmasters' intelligence estimates; and the Spot Pattern Test (with a reliability of 0.55) correlated at 0.76 and 0.75 with heads' intelligence estimates. The Dotting Test did best of all, correlating at about 0.8 with estimates of intelligence.

In summary, the results were supportive of the Binet approach; tests of higher mental functions appeared to be the most predictive of estimated intelligence. Of the twelve tests used by Burt (1909-10), six gave correlations of around 0.5 or below and six gave 0.5 or above. According to Burt,

The former six - the simple sensory and motor tests - are thus of little use in the empirical diagnosis of intelligence.

If Burt's (1909-10) study was a failure of the simple sensory approach to intelligence, as Burt himself considered it to be, it was only a relative failure. Burt's correlations replicated the modest correlations found between simple sensory measures in the visual and auditory modalities and estimates of intelligence that had been reported by Spearman (1904). Burt's hierarchy of correlations was not quite as perfect as that of Spearman but, he reckoned, it did not contain more deviations than would be expected by chance. With two authoritative studies and nothing of substance to contradict them Burt (1909-10) expressed confidence in the notion of *g*:

The main significance of this hierarchy of experimental performances, is, as it appears to me, that we are led to infer that all the functions of the human mind, the simplest and most complicated alike, are probably processes within a single system. A process typical of higher psychophysical 'levels' may be connected with a process typical of lower psychophysical 'levels'. Yet, this relatively small correlation is not a disproof, but a consequence of, their inclusive organisation within a single integrative system of psychical dispositions or neural arcs. The contrary assumption of a radical dichotomy between "the general mammalian foundation of the central nervous system" and the "specifically human capacity" of General Intelligence - towards which Dr Archdall Reid, and even Professor Thorndike seem to incline, - proves a serious barrier to the advance of the biological standpoint in individual psychology.

This early conceptualisation of a hierarchy of abilities, with basic psychological components at lower levels, was a presage of a similar model formulated by Jensen (1985b).

### **3.4 Intelligence and sensory discrimination after Spearman and Burt**

The studies discussed above were examined at some length because they provided sufficiently detailed methodology and analyses to make their conclusions supportable. Also, they corroborated each other without being straight replications: Burt's (1909-10) conclusion that simple sensory measures were useless in the diagnosis of intelligence does not detract from the fact that he, too, found a relation between school performance and

teachers' estimates of intelligence and the ability to make simple sensory discriminations, especially those involving pitch differences. These reports suggested that the correlation between intelligence and auditory (pitch) and visual discrimination was around 0.3. This is modest, but significant, and it is of considerable interest because of the apparent differences in complexity between the tasks.

Recall that Spearman (1904) advocated throwing out all studies prior to his own due to poor methodology and to their lack of statistical examination, and that of two studies re-examined by Spearman (1904), Seashore's data revealed a correlation of 0.24 between sensory acuity and imputed intelligence. Thorndike, Lay and Dean (1909) tested 37 female school students and 25 third year high school boys for ability to match to lines of different length and to boxes of different weights. These sensory discrimination task variables were correlated with estimates of intelligence and scholarship. Thorndike, Lay and Dean (1909), in the discussion of their results, were replying specifically to Spearman's (1904) claim that there was an *identity* between general discrimination and general intelligence, and they did not spend much time discussing the fact that they also had found a modest positive relationship between the two. Thorndike, Lay and Dean (1909) estimated that the true correlation between general discrimination and general intelligence was 0.23. Burt (1909-10) was more concerned with other matters in his own paper to linger on the fact that he, Spearman and Seashore were in agreement that there was some consistency across various studies in demonstrating a modest positive correlation between sensory discrimination and estimates of intelligence. In fact, the result was passed over in the consideration of more theoretical matters, which arose out of extreme positions. It is interesting to note that Thorndike, Lay and Dean (1909) did appear to accept the existence of general intelligence,

With young children a test designed to measure sensory discrimination may easily become, to a considerable degree, a measure of ability to understand instructions, that is, of one feature of general intelligence.

But this apparent sympathy with *g* lasted hardly more than a paragraph before a more basic disagreement emerged; Spearman's (1904) claim about the identity of general discrimination and general intelligence was criticised by Thorndike, Lay and Dean (1909), and there was no evidence of an awareness of the fact that both studies had found modest correlations that might point to the same conclusion,

The theoretical importance of Spearman's conclusion lies in the support which it would give, if verified, to the hypothesis that the efficiency of what may be called the general mammalian foundation of the nervous system is closely correlated with what may be called the specifically human neurone-connections. The present results support the contrary hypothesis, that the efficiency of a man's equipment for the specifically human task of managing ideas is only loosely correlated with the efficiency of the simpler sensori-motor apparatus which he possesses in common with other species.

In general there is evidence of a complex set of bonds between the psychological equivalents of both what we call the formal side of thought and what we call its content, so that one is almost tempted to replace Spearman's statement by the equally extravagant one that there is *nothing whatsoever* common to all mental functions, or to any half of them. [*Italics in the original.*]

The temptation articulated by Thorndike, Lay and Dean (1909) proved irresistible to many others and those, like Thomson (1939), who took this hypothesis seriously ignored the evidence that had prevented Thorndike's biting the apple.

There are two further reports, supervised by Spearman at University College, London, that are even less cited than those of Burt (1909-10) and Spearman (1904) in the accounts which have given a verdict on the intelligence/simple discrimination literature. These studies provided larger subject samples tested on more valid batteries of mental tests, and they calculated reliability coefficients for each test. Abelson's paper (1911) is introduced with the comment that there was little good evidence for or against the existence of *g*. Abelson's concern was that, while Binet and Simon's test was very useful and had, by that time, convinced the sceptical that mental ability may be measured,

They do not know what these tests measure or signify. The tests are isolated from the main body of scientific psychology. They neither derive much light from it, nor do they import much to it.

The stated purpose of Abelson's (1911) study was to discover the psychological principles underlying the tests. Eighty-eight girls and 43 boys were tested. These were "backward" children who were pupils in eight London County Council schools for "mentally defectives." The brightest children in these schools were selected for testing, i.e. those who were educable, had no sensory deficits and who often returned to normal schools within a few years. The tests were performed in a quiet room, every test being done at least twice. The entire study took three years to perform and involved eighteen visits to each child. Some of the tests (including, unfortunately for this review, pitch and weight discrimination) were considered to be too difficult for these children to perform, and the tests included in the final analyses were,

Tapping - using a "pointed instrument" to make as many taps as possible inside a square of paper. At each visit three ten second sessions were allowed and the first of these was discarded.

Crossing out an irregular line of rings.

Crossing out groups of four dots from rows that contained groups of three, four and five dots.

Memory tests for sentences, names and commissions. The last of these involved a series of increasingly complex instructions to be carried out by the child, e.g., "Put the matchbox on the armchair."

Discriminating the longer of two vertical lines.

Geometrical figures: this involved pointing to a point in a geometrical construction

that satisfies a command like, "Point inside the two circles and the triangle but only in one square." Items similar to this form a subtest of the present-day Cattell Culture Fair Intelligence Test.

Interpretation of pictures: having checked that the children were familiar with the objects in a painting they were asked to interpret the work in terms of intentions, emotions, etc.

Measures of reading ability were obtained from the teachers, and an estimate of practical intelligence was provided by teachers after they had asked themselves, "Which of these children she would soonest trust on an errand requiring the sharpest intellect." The reliabilities of Abelson's (1911) tests were high, with an average test-retest reliability of 0.83 and a range of 0.70 to 0.97. The correlation matrices for boys and girls were presented separately (and are reproduced in Appendices 1.1 and 1.2, respectively, toward the end of this thesis). The average inter-test correlation was 0.32 for girls and 0.26 for boys. There were a very few, near-zero negative correlations. The average correlation when tests were correlated singly with global rankings of intelligence for girls was 0.50 and for boys was 0.41. For the simple lines discrimination task the average correlation with the other eleven test scores was 0.30 (range 0.20 to 0.43) for girls and was 0.28 (range 0.11 to 0.47) for boys, i.e. brighter children made finer discriminations. Abelson (1911) estimated that any single test would correlate with the *g* factor at 0.57 for girls and at 0.51 for boys.

All of the above correlations were derived without correction for test-retest reliability which, although generally very high, was less than unity and, as a result, all correlations were depressed. The correlations were depressed further by the restricted range of ability in the sample. Having chosen children who were not educable at normal schools, but were the brightest at these special schools, Abelson (1911) probably had a group in the IQ range



from about 65 to about 85. The results offer three findings of interest. First, the finding of a correlation of about 0.3 between visual discrimination of line lengths and a wide range of mental tests was corroborated. Second, discrimination of line lengths correlated with Abelson's imputed *g* factor at about the same level as any test of the higher mental functions. Third, the correlation matrix, even in this relatively restricted sample, was almost universally positive and supported the supposition that there is a *g* factor underlying mental test performance.

### **3.4.1 Re-analysis of Abelson's (1911) data**

Abelson's (1911) final correlation tables (Appendices 1.1 and 1.2) were submitted to principal components analyses by the present author. For these purposes there was a rather small number of subjects, given the number of variables. The sample of girls was more adequate than the boys, having roughly 7.5 subjects for every variable. Using Kaiser's (1962) criterion, only those factors with latent roots greater than one were extracted. Principal components analysis of the boys' ( $n=43$ ) correlation matrix (Table 3.1) revealed a general factor, with most tests having substantial loadings on it (loadings ranged from 0.35 to 0.70, mean loading=0.57). In this sample 71.7 % of the total variance was extracted by four factors. The general factor accounted for 33.6% of the variance. In the girls' sample ( $n=88$ ) 67.1% of the total variance was extracted in four factors. The first factor (Table 3.2) was a general factor, accounting for 37.7% of the total variance and all tests were substantially positively loaded on it (loadings range from 0.46 to 0.76, mean loading=0.61). The discrimination task loaded at a very similar level on the general factor extracted from the boys' matrix (0.57) to that of the same test done on the girls' sample (0.58).

Varimax rotation of the components was carried out for both samples. The results are presented in Tables 3.1 and 3.2 for the samples of boys and girls, respectively. This procedure distributes the same percentage of the variance as principal components analysis,

**Table 3.1**

First principal component and varimax rotation results from principal components analysis of the matrix of correlations for boys (n=43) from the data of Abelson (1911). 71.7% of the common variance was extracted by 4 components.

	First Principal Component	Varimax Rotation Components			
		1	2	3	4
% Variance explained	33.6	20.5	16.4	18.1	16.7
Reading ability	0.70	0.05	0.22	-0.85	-0.25
Practical intelligence	0.69	0.11	0.01	-0.67	-0.49
Memory for sentences	0.61	0.83	0.15	-0.11	-0.06
Geometrical figures	0.61	0.68	-0.03	-0.16	-0.30
Crossing out dots	0.57	0.14	0.82	-0.32	-0.08
Discrimination of lines	0.57	0.17	0.36	-0.14	-0.54
Interpreting pictures	0.57	0.42	0.19	-0.11	-0.42
Arithmetical ability	0.61	0.32	0.10	-0.89	0.20
Memory for commissions	0.58	0.43	-0.27	-0.15	-0.68
Tapping	0.47	-0.07	0.38	-0.00	-0.77
Memory for object names	0.53	0.88	-0.12	-0.14	0.01
Crossing out rings	0.35	-0.11	0.91	-0.02	-0.20

**Table 3.2**

First principal component and varimax rotation results from principal components analysis of the matrix of correlations for girls (n=88) from the data of Abelson (1911). 67.1% of the common variance was extracted by 4 components.

	First Principal Component	Varimax Rotation Components			
		1	2	3	4
%Variance explained	37.7	16.2	18.5	17.3	15.0
Practical intelligence	0.76	0.30	0.53	-0.44	0.23
Memory for commissions	0.68	0.63	0.49	-0.32	-0.12
Tapping	0.63	0.08	0.23	-0.13	0.89
Crossing out rings	0.63	-0.12	0.43	-0.55	0.38
Memory for sentences	0.64	0.81	0.02	-0.30	0.16
Interpreting pictures	0.64	0.45	-0.06	-0.25	0.71
Arithmetical ability	0.63	0.14	0.67	-0.25	0.17
Geometrical figures	0.61	0.28	-0.05	-0.75	0.22
Discrimination of lines	0.58	0.23	0.37	-0.23	0.33
Crossing out dots	0.56	0.06	0.19	-0.77	0.05
Memory for object names	0.50	0.64	0.26	0.17	0.29
Reading ability	0.46	0.07	0.87	0.08	0.03

but attempts to minimise the number of tests loading on any one factor while ensuring that significant loadings are as high as possible. The first rotated factor was very similar for both samples, with high loadings on all three memory tests, the geometrical figures (which involves memory) and the interpretation of pictures. This appeared to be a memory factor. The second factor for the girls had highest loadings for arithmetic, reading and imputed intelligence and was a school intelligence/educational factor. In fact, this was very similar to the third factor in the boys' sample where an even clearer educational factor emerged with the same three main loadings. The third factor in the girls' sample had loadings for, mainly, crossing out rings and dots and the geometrical figures test and may have been a perceptual-motor factor. Again, the boys had a very similar, but clearer, factor two where the main loadings were for the crossing out dots and crossing out rings tasks. In both samples the fourth factor loaded highly on tapping and interpretation of pictures but other loadings did not correspond so well.

Given the small samples (in factor analytic terms) and the restricted range in the samples there was remarkable agreement between the results of the two groups. The coefficient of congruence comparing the first unrotated principal component for the sample of boys and girls was 0.98. Abelson (1911) concluded that,

[the]...idea that the tests are mere laboratory artefacts, having no relation to ordinary life, falls to the ground.

The results of this study, which showed considerable cross-sample validity in this re-analysis are rarely if ever alluded to in the recent accounts of the Binet approach versus the simple sensory measures approach to intelligence. Whereas the impression that many of the above writers gave was of one approach's success and another's failure, a truer picture might be of the irresistibility of a practical measure of mental ability which cut correlational psychology adrift from a better understanding of the basic processes involved in intelligent functioning, leading to the state of affairs discussed in Chapter 1. Abelson's

(1911) study provided a good illustration of the difficulty of carrying out discrimination tests (two of them had to be omitted from the above study), and Burt (1909-10) had discovered previously how unwieldy they were. These were some of the last few voices at the shore, warning that the drift of the mental testers from the *terra firma* of experimental psychology would do more harm than good. The simple sensory approach was not a failure. Modest and surprising results were obtained, but they were outshone by the needs of the educators and testers, and by the debates of the advocates of *g* versus those, like Thomson (1939), who envisaged human cognitive ability as a number of separate abilities.

### **3.4.2 Re-analysis of the studies by Carey (1914-15 and 1915-17)**

The last study that will be mentioned in detail from the same period is that of Carey (1914-15, 1915-17). In a series of reports prepared from his D.Sc. dissertation, Carey examined the correlations between intelligence estimates and sensory and memory measures for schoolchildren. Over 150 children from London County Council Elementary Schools were tested. Estimates of school intelligence, practical intelligence, painstaking and social status were obtained from teachers. Using the pooled results of at least two separate tests Carey obtained visual, auditory and tactile discrimination measures. Visual, auditory and verbal memory tests were given. Global exam marks were obtained and children were tested for their ability to give the opposite of a word and to unscramble sentences. Finally, tests called holes and drawings were administered. Again, Carey gave reliability estimates for all of the tests, including the estimates of intelligence (although informants tended to know each other's opinions). The mean reliability was 0.70 (range 0.53 to 0.86).

Carey's (1915-17) final (uncorrected) correlation matrix is shown in Appendix 1.3. The mean correlation of all tests with visual discrimination was 0.28 (range 0.00 to 0.51). The mean correlation of all of the other tests with auditory discrimination was 0.22 (range -0.01 to 0.46). Tactile discrimination did more poorly and correlated at only 0.02 on



average. These replicated the earlier results of Seashore, Spearman (1904), Thorndike, Lay and Dean (1909), Burt (1909-10) and Abelson (1911), in that there was a significant, if modest in extent, correlation between tests of auditory and visual discrimination and other mental tests. Table 3.3 shows the results of the principal components analysis for the Carey matrix performed for the present review ( $n > 150$ ). Six components accounted for 76.4% of the total variance, but only the first component accounted for significantly more than 10% of the variance. Again this was a general factor with highly significant loadings for all tests, except tactile discrimination. Visual and auditory discrimination loaded at 0.57 and 0.45, respectively, on the general factor. Varimax rotation of this matrix (Table 3.3) revealed four factors that each accounted for more than 10% of the total variance. The first had high loadings for exam marks, sentences test, school intelligence, practical intelligence and opposites. This appeared to be a verbal/educational factor. The second factor was a memory and visual discrimination factor. The third factor loaded almost exclusively on the drawings and holes tests. Factors four and six were related to auditory and visual discrimination while factor five had high loadings for social status, painstaking and school intelligence.

### 3.5 Conclusions

The main conclusion from this review and re-analysis is that there has been some consistency in finding that sensory discrimination estimates, especially pitch discrimination, correlated at modest levels with tests or estimates of mental ability. The reported history of this research is often a less than accurate representation of the findings of the original studies. Two much-cited and influential studies, those of Sharp (1898-9) and Wissler (1901), contain deficiencies that render them of little use in reckoning whether the ability to make fine discriminations is associated with higher cognitive abilities. A series of less-cited studies was presented, and in some cases re-analysed, and they appeared to be consistent in finding modest associations between sensory discrimination indices and tests of mental ability. However, the interpretation of these results is not



**Table 3.3**

First principal component and varimax rotation results from principal components analysis of the matrix of correlations for all subjects ( $n > 150$ ) from the data of Carey (1915-17). 76.4% of the common variance was extracted by 6 components.

	First Principal Component	Varimax Rotation Components					
		1	2	3	4	5	6
%Variance explained	34.9	21.8	15.8	10.6	7.6	12.3	8.4
School intelligence	0.87	0.75	0.20	0.12	0.12	-0.51	0.03
Practical intelligence	0.63	0.70	0.07	0.11	-0.20	-0.18	-0.03
Painstaking	0.70	0.39	0.22	0.22	0.14	-0.70	-0.19
Social status	0.52	0.10	0.03	0.17	0.02	-0.87	0.24
Visual discrimination	0.57	0.19	0.67	0.26	-0.02	-0.06	0.09
Auditory discrimination	0.45	0.19	0.23	0.11	-0.03	-0.14	0.80
Tactile discrimination	0.06	-0.00	0.08	-0.05	0.92	-0.10	-0.00
Visual memory	0.48	0.14	0.82	0.08	0.11	0.10	-0.05
Auditory memory	0.53	0.04	0.73	-0.07	0.01	-0.32	0.24
Verbal memory	0.49	0.33	0.45	-0.10	-0.34	-0.29	-0.44
Exam marks	0.75	0.90	0.06	0.26	0.19	-0.07	-0.07
Opposites	0.78	0.61	0.50	-0.04	-0.12	-0.13	0.40
Sentences	0.66	0.77	0.21	-0.01	-0.08	-0.04	0.21
Drawings	0.56	0.26	0.26	0.83	0.07	-0.07	-0.07
Holes	0.35	0.04	-0.04	0.79	-0.11	-0.22	0.22

unproblematic. For instance, Abelson's study of mildly mentally handicapped individuals is not necessarily applicable to the general population and, for these subjects especially, Thorndike, Lay and Dean's (1909) caution that, in samples of children, most so-called 'simple sensory' tests may be construed as tests of instruction-following tests might have considerable force.

Such caveats notwithstanding, a full account of any basic processes involved in tests which purport to estimate individual differences in human intelligence will have to account for these results and the results of more recent studies (discussed in Chapter 2) that have suggested that pitch discrimination is associated with psychometric measures of intelligence. This historical review and re-analysis reconfirms the need to take seriously the suggestions made by recent authors that pitch discrimination might be involved in the correlations between cognitive ability test level and auditory inspection time task ability.

## **Chapter Four**

### **An improved auditory inspection time task: associations with cognitive ability and visual inspection time**

#### **4.1 Introduction**

The aims of this Chapter were outlined at the end of Chapter 2. The first aim was to design a modified auditory inspection time (AIT) task and to construct a device to generate the stimuli. The second aim was to examine the performance curves obtained by subjects on the modified AIT task as the stimulus duration was varied. The third aim was to determine the degree of correlation between AIT threshold estimates and visual IT thresholds. The fourth aim was to determine the degree of correlation between AIT thresholds and verbal and non-verbal cognitive ability test scores.

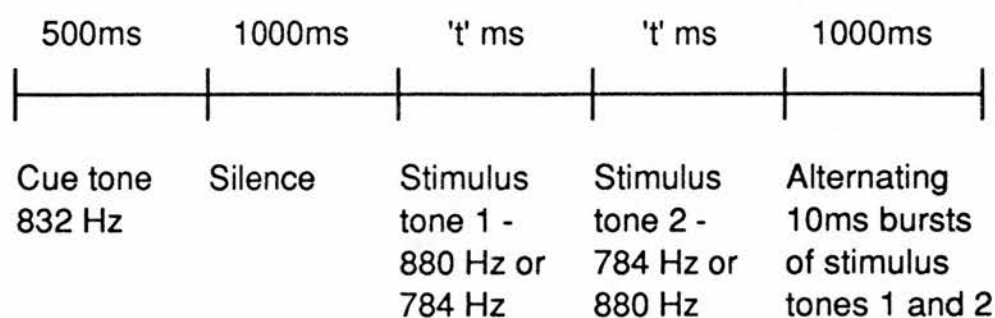
##### **4.1.1 Development of a modified AIT task**

Following the discussion in Chapter 2 on the limitations of earlier attempts to measure inspection times in the auditory modality, an attempt was made to design a task which would overcome the main problems encountered with previous AIT tasks. It was decided to retain approximately the same pitch difference between the stimulus tones as was used by Brand and Deary (1982) and by Nettelbeck, Vreugdenhil and Edwards (1986), i.e. about 100 Hz, because this degree of separation is standard in much work in ability testing in auditory psychophysics (Massaro, 1970; Raz, Willerman, Ingmundson and Hanlon, 1983). Further, the adoption of an auditory backward mask slightly modified from that used by Nettelbeck, Edwards and Vreugdenhil (1986), and consisting of alternating short bursts of the two stimulus tones, appeared to offer the possibility of obtaining inspection times which were not too brief and, therefore, were not prone to the problems met with very low duration stimuli, as discussed by Irwin (1984).

Therefore, the task which was constructed had the following design. Subjects were to hear an auditory cue tone, which was midway between the two stimulus tones in frequency. The offset of the cue tone was to precede the first of the two stimulus tones by 1000ms. After this 1000ms of silence the two stimulus tones were played, each for an identical period of time ( $t$ ), one immediately after the other, i.e. there was to be no interstimulus interval in the sense used by Raz, Willerman, Ingmundson and Hanlon (1983). The tone envelopes were to have instantaneous rise and fall times. Immediately after the offset of the second tone, the backward mask, consisting of alternating 10ms bursts of the two stimulus tones, was to be played for 1000ms. The parameter to be manipulated to estimate the subject's inspection time was the duration of the stimulus tones (time  $t$ ) and the outcome variable was the correctness of the subject's response to the question, "was the order of the tones 'low-high' or high-low?".

Given these specifications an apparatus was constructed in-house by the Electronics Workshop, Department of Psychology, University of Edinburgh under the direction of the late Mr D. Wight, to produce the relevant sounds. Figure 4.1 gives a schematic account of the temporal sequence of each trial and Appendix 2 shows the circuit diagram for the apparatus designed and constructed for the studies to be reported in this thesis. Stimuli were produced using an XR320 monolithic timing circuit. The auditory mask consisted of rapidly alternating bursts of both tones (10ms each) provided by a multivibrator circuit. The stimulus unit was interfaced with a BBC B microcomputer which controlled the duration of cue, stimuli, mask and intervals between these elements.

Once the device was constructed, pilot testing to determine the range of stimulus tones took place on the author, the technicians involved in constructing the apparatus, and on 15 Youth Training Scheme workers who were undergoing visual inspection time and mental testing in the Department of Psychology as part of another research project (Egan, 1991).



**Figure 4.1**

Schematic representation of the temporal order of events in each item of the modified auditory inspection time task, not drawn to scale. Stimulus tones lasted for identical times - 't' ms - ranging from 6ms to 200ms. Subjects made responses at leisure by indicating the order - 'High-Low' or 'Low-High' - of the two stimulus tones. There was no stimulus-mask gap.

This involved using the PEST adaptive staircase procedure (Taylor and Creelman, 1967; this technique is described more fully below) to estimate inspection times using an 85% threshold. As a result of this preliminary testing, it was found that no-one obtained an inspection time briefer than 20ms, and that no-one had an inspection time longer than 140ms. Seven of the 15 Youth Training Scheme subjects (46.7%) were unable to perform the discrimination at any duration up to 240ms, an aspect of auditory discrimination tasks noted in previous studies (Brand and Deary, 1982; Nettelbeck Edwards and Vreugdenhil, 1986) and representing a proportion close to that reported by Massaro for subjects unable to achieve near-perfect asymptotic performance in a different auditory psychophysical procedure using tones of approximately the same pitches as those used here (Massaro, 1976; Kallman and Massaro, 1979).

Subsequently, in order to collect data on the performance curves of subjects tested on the modified AIT task, a version of the auditory inspection time test was created which presented each subject with 10 trials at each of 13 durations, i.e. the psychophysical method of constant stimulus durations (descending series) was used. Because adaptive staircase procedures present stimuli in response to the performance of the individual subject, this method was not suitable for collecting data to examine systematically the influence of stimulus duration on the correctness of the discrimination. As a result of the pilot testing, the durations used (in milliseconds) were 200, 150, 125, 100, 85, 70, 55, 40, 30, 20, 15, 11, 6. As stated above, the trials were presented in decreasing order of duration, with the ten trials at 200ms given first and the ten trials at 6ms given last.

To precede the test stimuli, a detailed and standardised set of instructions was created. These test instructions introduced subjects to each component of the AIT task items separately, i.e. the cue, the tone pairs and the mask, and they built up AIT task items component by component in order to allow subjects to become familiar with the task demands. Before the 130 trials in the task proper, there were 19 practice trials. The first

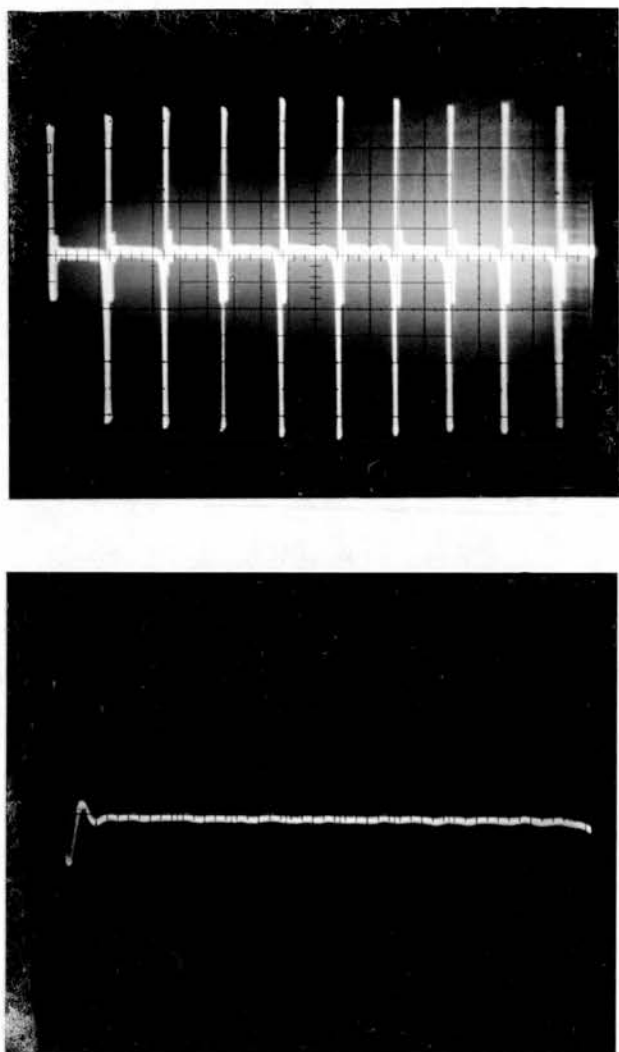


ten practice trials presented subjects with the cue tone followed by two stimulus tones at a duration of 200ms, but *without a backward mask*. These trials were included to detect those subjects who were having problems with the discrimination required to perform the task. Following these trials came nine trials with stimulus tone durations of 200ms. The offset of each of these tone pairs was followed by the backward mask. The entire test, comprising the detailed spoken test instructions, the 10 unmasked practice trials, the nine masked practice trials and the 130 test trials, was recorded on a UHER reel-to-reel tape recorder at high speed (7.5 inches per second). This ensured, as far as was possible, that groups of subjects would be presented with a standard, unchanging task. High quality tape recordings are commonly used to present other auditory tests (Raz, Willerman, Ingmundson and Hanlon, 1983; Watson, Johnson, Lehman, Kelly and Jensen, 1982; Bentley, 1963).

#### **4.1.2 Physical characteristics of the test stimuli**

Having compiled and recorded the modified AIT test, various checks were undertaken to ascertain whether the stimuli were reproduced faithfully from the apparatus. The first check was to assess whether the tape-recorded test was accurately reproducing the durations and frequencies which had been generated by the original apparatus. The output of the recording of the AIT test played on a UHER reel-to-reel tape recorder was examined on an oscilloscope. Measurements of waveform peak-to-peak distances established that stimulus tone durations the frequency differences between the tones were reproduced accurately. Figure 4.2 was photographed from the oscilloscope display of a 6ms item and demonstrates that the stimulus tone has an instantaneous rise time; each wave is clearly visible and there is no detectable noise on the output of the tape recording.

None of the subjects in the pilot testing could discriminate AIT task stimulus tones played for less than about 20ms, which meant that no check on the accuracy of the response correctness could be made by listening to the tones. Moreover, because it appeared likely



**Figure 4.2**

Photographs of the oscilloscope traces of the test item output from the tape recorded AIT test. Both pictures represent the waveforms of a 6ms item, and in both pictures the oscilloscope trace was triggered by the onset of the first stimulus tone. In the upper photograph, each large oscilloscope screen section square represents about 1ms. Therefore, the trace may be measured to show that the first six peak-to-peak distances are wider than the later distances, i.e. the item is 'Low-High'. In the lower photograph each large square on the oscilloscope screen represents about 100 microseconds. Therefore, this trace shows the near-instantaneous onset of the first wave of the first stimulus tone.

that the very brief tones might suffer worse had there been significant loss of quality in the recording process, the Senior Audio-Visual technician in the Department of Psychology (Mr J. Cuthbert) was asked to make the 'High-Low' versus 'Low-High' discriminations of the test items with durations of 15, 11 and 6ms (30 in all) using only their displays on the oscilloscope. This was done by measuring the peak-to-peak distances after stimulus tone onset and comparing those with the peak-to-peak distances after the nominal duration of the first and second tones ('t' ms). Using this procedure all of the items at the three briefest durations were correctly classified as 'high-low' or 'low-high' pairs.

#### **4.1.3 Aims of the study**

To recap, the remaining aims of the study conducted in this Chapter were as follows. First, the subject performance characteristics of the auditory inspection time test were examined. If the AIT test behaves like the visual inspection time test one would expect that, as stimulus presentation time increases, the accuracy of the discrimination would increase in the form of a cumulative normal ogive as described by Vickers, Nettelbeck and Willson (1972).

The second aim of the study was to determine the degree of correlation between the auditory inspection time test and three forms of the visual inspection time test which have been found to correlate significantly with IQ-type tests, *viz.* the standard two vertical lines test (Vickers, Nettelbeck and Willson, 1972; Nettelbeck and Lally, 1976), the horizontal lines version of the test (Mackenzie and Bingham, 1985; Mackenzie and Cumming, 1986) and the 'Longstreth' IT test which used 'o' and '/' characters as stimuli and a combination of the two stimuli as the backward mask (Longstreth, Walsh, Alcorn, Szeszulski and Manis, 1986).

The final aim of the study was to determine the degree of correlation between the auditory inspection time test and verbal and non-verbal cognitive ability tests.

## **4.2 Methods**

### **4.2.1 Subjects**

Subjects were 120 second year psychology students studying in the Department of Psychology, University of Edinburgh. There were 81 women and 39 men. No sex differences have been found in IT studies, and sex differences were not examined in this study. The mean age of the subjects was 20.39 years (S.D. 2.93). Subjects' first language was English and only those with adequate corrected or uncorrected vision were used in the IT analyses. Subjects reporting hearing difficulties were not included in the study, but hearing ability was not tested formally. The limited range of cognitive ability among undergraduates is likely to attenuate the IT-IQ correlation. Nevertheless, this population continues to feature prominently in IT studies, probably because: they are readily available to take part in lengthy studies involving psychometric and psychophysical tests; they are relatively relaxed in the laboratory setting; and they are generally free from the problems associated with ageing and poor health that are likely to be confounding variables in such studies.

### **4.2.2 Psychometric Tests**

#### **Advanced Progressive Matrices**

The Raven Advanced Progressive Matrices test, 1962 Revision (Raven, Court and Raven, 1977) was selected to test non-verbal cognitive ability because it was specially designed to discriminate among high ability subjects. Both sets of the test were used. Set I, which was not scored, has 12 items that served to familiarise the subjects with the form of the items. Set II was administered as a timed (40 minute) test. It has 36 items of increasing difficulty. The test involved the subjects in scrutinising a pattern comprising abstract shapes with a logical relationship between the elements of the pattern. Each pattern has a piece missing and subjects decide which of several alternatives correctly completes the pattern.

### **Mill Hill Vocabulary**

The Mill Hill Vocabulary Form I Senior - Synonyms (Raven, Court and Raven, 1977) provided a test of vocabulary level and, although it was intended for a general adult population, it represented the more difficult form of the Mill Hill Vocabulary test. The test involves reading a word and deciding, by underlining, which of six alternatives provides a synonym for the target word.

### **Alice Heim 5**

The Alice Heim 5 (Heim 1968) test is a two part ability test which was designed to discriminate among high ability subjects. Part A is verbal-numerical and, in the course of 36 items, involved the subject in solving series, classification and other reasoning problems with words and numbers. Part B is non-verbal and involved the subjects in solving reasoning problems with geometric shapes. Each section has a time limit of 20 minutes.

### **4.2.3 Inspection Time Tests**

#### **Visual inspection time tests**

Three visual IT tests were used for this study. Stimuli were presented on a microcomputer screen. The psychophysical method used to estimate visual IT thresholds was a programmed version of the PEST adaptive staircase method (Taylor and Creelman, 1967). At the end of the staircase this procedure provided an IT estimate for a subject at the 85% level of discrimination accuracy. The starting stimulus duration for the PEST algorithm was 200ms. The stimulus duration change, i.e. each step in the staircase, was halved in size with each reversal of the staircase. The test was stopped when the algorithm attempted to change from 2 to 1 ms. The rules for changing duration were as follows. The initial change of duration in the staircase was 75ms. If a subject completed five consecutive trials at a given stimulus duration correctly, the duration of the next five stimuli was reduced. If more than one error was committed in a block of five trials, the stimulus duration was



increased for the next block of five trials. If one error was made in a block of five trials, more trials were given. If an error was committed in any of the next five trials, the duration was increased but, if no further errors were made at that duration, i.e. if the subject then had nine out of 10 correct, the next stimulus duration was reduced. The subject's inspection time represented the algorithm's estimate of the stimulus duration at which the subject was 85% correct in his or her responses. The PEST algorithm has been found to produce reliable discrimination thresholds, and an attractive feature of the procedure is that it has been found to give thresholds for naive subjects in some discrimination tasks that are close to those obtained from practiced subjects (Stillman, 1989).

Easy items at the beginning of each IT test provided practice and served to familiarise the subject with the requirements of the tests. In each of the visual tasks there was a rest after 10 trials. The subject initiated the restart. If the subject had performed at 80% or better in the last 5 items the message 'Well Done' appeared on the screen. If the subject had scored less than 80% correct in the last 5 items the message 'Pay Close Attention' appeared. All stimuli, screen backgrounds and general background light levels were checked for equality using a light meter. Stimuli were presented using a screen luminance of 2800 lux ca. on a background of 44 lux ca. Monitor screen controls were made inaccessible to the subjects.

All three visual tests were designed to be self-administered and self-paced. Trials in each of the three tests had a similar sequence. A visual cue, which was identical to the backward mask in form and location, lasted 500ms. Cue offset preceded the stimulus onset by 1000ms. Immediately after the stimulus offset the backward mask was presented and lasted for 1000ms. Subjects responded at their leisure, with instructions to maximise accuracy rather than speed of response. A response initiated the onset of the next item.

Vertical Lines Test. The first visual IT test used was a form of the standard vertical lines



IT stimulus. The long line was 5cm and the short line was 3cm in length. The two lines were 2cm apart, and they were approximately 2mm thick. The mask comprised a pair of identical vertical lines, 7cm long and 5mm wide, entirely surrounding the area of the stimulus lines. Subjects were required to indicate, by pressing one of two computer keys, whether the long line (stimulus presentations always included one long line and one short line) was on the left or on the right. In this test, as with all other IT tests described in this thesis, a record was taken of the correctness of each response, and no record was kept of the response time, since it was emphasised that subjects should respond slowly and with maximum accuracy.

Horizontal Lines Test. This was a modification of the test used by Mackenzie and Bingham (1985). The stimuli were two horizontal lines of the same lengths as the vertical lines. The masking lines were 7 cm long. The lines were about 1mm thick, as were the masking lines. Unlike the Mackenzie and Bingham stimuli, the lines were presented in the same horizontal plane and they always appeared in the same area of the screen. There appeared to be no advantage in following the practice of Mackenzie and Bingham (1985) by having the stimuli appear at unpredictable locations on the screen. In all other extant visual inspection time tests the stimuli are presented at fixed locations, allowing subjects to fix their attention at an appropriate location and thereby preventing errors which might be caused by attending to the wrong area of the screen. Subjects' tasks and responses were essentially the same as for the vertical lines test.

'Longstreth' Task. This test was similar to that used by Longstreth, Walsh, Alcorn, Szeszulski and Manis (1986). The stimulus was either a diagonal slash about 1cm in length *or* a rectangle measuring 8mm by 6mm. The mask was a combination of the two stimuli that occupied exactly the same area of the screen. Subjects were asked to indicate, by pressing one of two keys, whether they had seen the diagonal slash *or* the rectangle prior to the mask. Note that this IT test, unlike the previous two, did not involve the

comparison of two simultaneously presented stimuli.

### **Auditory inspection time**

This test was described above. Briefly, it was a fixed-paced temporal order pitch discrimination test using the method of constant stimulus durations (descending series). Stimulus tones were 880 Hz (high) and 784 Hz (low). Test stimuli were recorded and then presented to subjects using a UHER reel-to-reel tape recorder run at high speed. The output from the tape recorder was relayed via an auditory network to headphone sets connected to sockets in individual quiet basement cubicles. Sound levels were equalised for all headsets at 80 dB for cue, stimuli and mask. Each auditory trial consisted of a cue tone (832 Hz) lasting 500ms, 1000ms of silence, a stimulus pair ('high-low' or 'low-high') of tones of given identical durations, and the 'warble' mask for 1000ms. There was no gap between the second stimulus tone and the mask. About 8 seconds of silence between items gave subjects adequate time to tick 'high-low' or 'low-high' on a response sheet.

Detailed instructions for this test were recorded on the pre-recorded test tape. Two practice blocks, of ten and nine trials, respectively, followed. These provided an introduction to the task at easy stimulus durations (200ms), and the first block of ten trials was given without an auditory mask. The experimental session consisted of thirteen blocks of ten stimulus pairs presented at decreasing durations (in milliseconds), i.e. 200, 150, 125, 100, 85, 70, 55, 40, 30, 20, 15, 10 and 6. Subjects were required to score 90% or better in the second practice and first experimental blocks before being considered able to discriminate pitch. The auditory IT of a subject was reckoned to be the last duration at which the subject was 90% correct for that and all easier (longer) durations. Single errors at longer stimulus durations could be 'recouped' by 100% success at more difficult durations.

#### **4.2.4 Procedure**

Subjects were randomly allocated to two groups. Group 1 subjects took the psychometric tests before the IT session. Group 2 subjects completed the sessions in the reverse order. For each subject all tests were performed in a single afternoon session with the exception of AH5, which had been completed by a randomly chosen subset of 60 subjects in afternoon sessions 2 months previously.

The psychometric tests were administered as group tests under examination conditions with constant supervision, according to the instructions in the respective test manuals.

Before the inspection time testing took place, subjects were given a general briefing of the order of events, a general description of the visual tests and instructions on how to load and administer the tests for themselves. All subjects were familiar with the microcomputer network from previous practical sessions. For the AIT test, subjects, in groups of about 18, went to individual quiet cubicle rooms and put on a set of headphones. They listened to the pre-recorded auditory inspection time test tape which gave full instructions for test completion. The cue, stimuli and mask were introduced separately, and specimen items were built up sequentially. Several items were then played with the appropriate answers ("high-Low" or "low-high"). Practice blocks 1 and 2 and the 13 experimental blocks followed. There was a short rest pause between each block of ten trials. The appropriate answer to each item was endorsed by subjects on a response sheet. Total testing time for the auditory IT test was about 40 minutes.

After the auditory test the subjects undertook the visual inspection time tests. All subjects were instructed to perform the vertical lines test. Due to time constraints, subjects performed only one of the two other visual IT tests. Following the vertical lines test, after a short rest, half of the subjects performed the horizontal lines test and half of the subjects performed the Longstreth task.

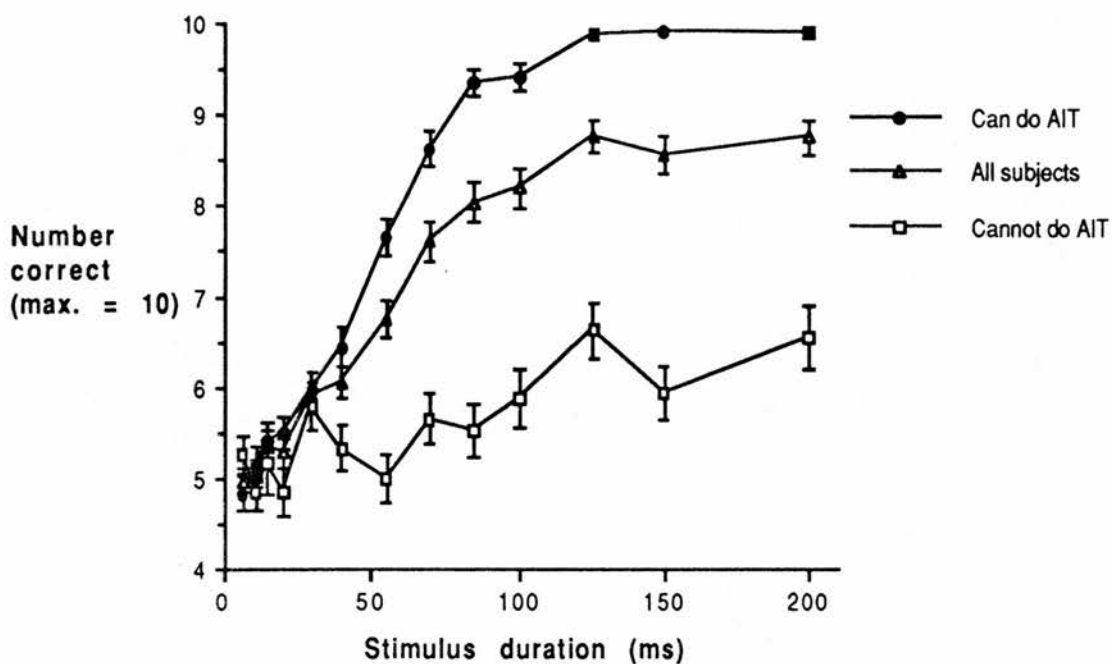
## **4.3 Results**

### **4.3.1 Psychophysical characteristics of the auditory inspection time test**

If longer durations of the test stimuli in the auditory inspection time test were providing more information for the decision processes than the briefer tones according to the hypotheses of the accumulator model of discrimination, then one should expect that the relationship between the duration of the stimulus tones and the number of correct discriminations would be described by a cumulative normal ogive (Vickers, Nettelbeck and Willson, 1972). To examine whether subjects' performances on the AIT task designed for the present study appeared to resemble this function, mean scores were calculated at each stimulus duration for 117 of the 119 subjects who provided full data for the auditory inspection time test (two subjects omitted a test item response at durations well below their inspection time).

The results are shown in figure 4.3. This shows that those subjects who were defined as being able to discriminate the test stimuli reliably at long durations, and who provided fully completed response sheets in the auditory inspection time task ( $n=78$ ), were scoring at near to chance levels for the shortest durations and were obtaining near-perfect scores at longer durations; i.e. the task showed neither ceiling nor floor effects for these subjects.

Moreover, there appeared to be a sufficient number of durations between chance and perfect responding to allow for a reasonable discrimination among subjects on the basis of their individual differences in the test. An asymptote at approximately 100% response accuracy occurred between 100 and 150ms. Figure 4.3 also shows the performance curves for all subjects in the study who returned complete response sheets ( $n=117$ ), and for the subjects who were reckoned not to have had sufficiently good pitch discrimination ability to allow them reliably to discriminate the test stimuli at any duration used in the present study ( $n=39$ ). As expected, all subjects scored at near to chance levels on the trials involving the briefest durations. For those subjects who could not perform the auditory



**Figure 4.3**

Means (standard errors) of the number of correct responses in the AIT task at 13 different durations for subjects classified as being able to perform the AIT task (closed circles;  $n=78$ ), all subjects in the study (open triangles;  $n=117$ ) and those subjects classified as not being able to perform the AIT task (open squares;  $n=39$ ).



discrimination reliably, the curve less clearly approximated to the form of a normal ogive, and the maximum score for these subjects was between 60-70% correct at between 100-150 ms. The fact that an asymptote appears to occur for these subjects, and the fact that it occurs at about the same duration in the groups who could and could not perform the discrimination involved in the test, might suggest that the latter group were performing at their maximum level of performance accuracy at about the same durations as the group who were more able in pitch discrimination.

#### **4.3.2 Cognitive ability and inspection time test scores**

All subjects provided Mill Hill Vocabulary (MHV) scores. Advanced Progressive Matrices (APM) scores were obtained from 119 of the 120 subjects due to a single incomplete answer sheet. Alice Heim 5 (AH5) scores were available for 60 subjects. After eliminating subjects who reported unsatisfactory vision and a few anonymous response sheets, 105 Vertical Lines inspection time test results were available, 51 subjects' results were collected for the Horizontal Lines IT test and 50 subjects provided data for the Longstreth IT test. Auditory IT response sheets were available for 119 subjects. Of these, 80 subjects met the criteria for AIT threshold estimation. Therefore, in this high ability group 32.8% of subjects found the discrimination (two tones separated by 96 Hz) too difficult to perform reliably at the longest durations involved in the present task. This might be caused by the fact that they were poor at pitch discrimination or because of their having auditory IT thresholds longer than 200ms.

Means and standard deviations for all measures are given in Table 4.1. The APM mean of 24.9 corresponded to an IQ mean of about 124 with a standard deviation (SD) of about 7 points. This was similar to, if slightly higher than, the previously reported means for undergraduates (Raven, Court and Raven, 1977). The MHV score of 31.9 (SD 3.46) was not directly comparable to population means as it was based on only the synonym test, but it also appeared to have a restricted range of scores. The AH5 test mean of 36.4 (total score for Parts A and B) and standard deviation of 7.45 were close to those reported on previous



**Table 4.1**

Means and standard deviations for the tests of cognitive ability and inspection time

	n	Mean	Standard Deviation
Advanced Progressive Matrices	119	24.9	3.46
Mill Hill Vocabulary	120	31.9	3.46
Alice Heim 5 (Verbal-Numerical)	60	16.9	3.78
Alice Heim 5 (Non-verbal)	60	19.6	4.76
Alice Heim 5 - Total Score	60	36.4	7.45
Vertical Lines IT	105	43.6ms	27.9
Horizontal Lines IT	51	75.4ms	29.2
Longstreth IT	50	29.8ms	29.3
Auditory IT	80	75.8ms	27.5

**Table 4.2**

Pearson's r correlations among different tests of cognitive abilities.

	AH5 Verbal- Numerical	AH5 Non- Verbal	AH5 Total	Mill-Hill Vocabulary	Advanced Progressive Matrices
AH5 Verbal-numerical	-	0.50*** (n=60)	0.83*** (n=60)	0.31* (n=60)	0.42*** (n=59)
AH5 Non-verbal		-	0.87*** (n=60)	0.46*** (n=60)	0.48*** (n=59)
AH5 Total			-	0.43*** (n=60)	0.52*** (n=59)
Mill Hill Vocabulary				-	0.25** (n=119)

\* p&lt;0.05, \*\* p&lt;0.01, p&lt;0.001

groups of university undergraduates (Heim, 1968). Therefore, the subject population performed at levels expected on the basis of other undergraduate samples, and had relatively low standard deviations on test scores.

Table 4.1 shows that all three visual tests had similar standard deviations. The vertical lines IT task proved to be relatively easy (mean 43.6ms, SD 27.9); 18 subjects (17.1%) scored at or better than the lower presentation limit (20ms) of the monitor screen. This ceiling effect occurred with twenty subjects (40%) in the Longstreth task (mean IT estimate=29.8ms, SD 29.3). The horizontal lines IT task had the highest mean value, at 75.4ms (SD 29.2) and only one subject did better than 20ms on this test. The mean auditory IT for those subjects classified as being able to perform the AIT task was 75.8ms (SD 27.5). Therefore, the auditory task had a similar standard deviation to the three visual tasks.

Table 4.2 contains the Pearson  $r$  correlations among the psychometric tests of cognitive ability. All correlations were positive and significant. The MHV test provided the two lowest correlations, 0.25 with the APM and 0.31 with the verbal-numerical section of the AH5 test. Ignoring the within-AH5 correlations the others fell into the range of 0.42 to 0.52.

Table 4.3 provides the correlations among the various tests of IT. All correlations were positive and all but one was significant. The standard vertical lines IT task correlated at 0.48 with the horizontal Lines IT test, at 0.39 with the Longstreth IT test and at 0.24 with the auditory IT test. The auditory IT test correlated at 0.20 with the horizontal Lines IT test and at 0.53 with the Longstreth IT test. Despite marked ceiling effects in the vertical lines and Longstreth IT tests, the correlations between horizontal lines and Longstreth and the classical vertical lines IT test were highly significant.

Table 4.4 shows the correlation coefficients between the four IT measures and the cognitive

**Table 4.3**Pearson's  $r$  correlations among the four tests of Inspection Time.

	Vertical Lines	Horizontal Lines	Longstreth	Auditory
Vertical Lines	-	0.48*** (n=51)	0.39** (n=46)	0.24* (n=68)
Horizontal Lines		-	Not Tested	0.20 (n=32)
Longstreth			-	0.53*** (n=33)

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ **Table 4.4**Pearson's  $r$  correlations between the tests of psychometric intelligence and Inspection Time

	AH5 Verbal- Numerical	AH5 Non- Verbal	AH5 Total	Mill Hill Vocabulary	Advanced Progressive Matrices
Vertical Lines	-0.31* (n=51)	-0.27+ (n=51)	-0.33* (n=51)	-0.02 (n=50)	-0.16 (n=104)
Horizontal Lines	-0.22 (n=20)	-0.32 (n=20)	-0.32 (n=20)	-0.17 (n=51)	-0.23 (n=50)
Longstreth	-0.06 (n=32)	-0.41* (n=32)	-0.29+ (n=32)	-0.31* (n=50)	-0.28* (n=59)
Auditory	-0.11 (n=40)	-0.40** (n=40)	-0.31* (n=40)	-0.27* (n=80)	0.05 (n=80)

+  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$

ability tests. Only one correlation was not in the expected direction and this, even when corrected for attenuation of the ability range (Table 4.5), was near to zero. The ability test with the largest standard deviation and, by implication, the best discriminating power achieved the most consistent results. The AH5 total correlated with all four IT measures at very similar levels, between -0.29 and -0.33; two of these were significant and one was a trend. The AH5 non-verbal subtest produced a similar range of correlations and exceeded the AH5 total with the auditory and Longstreth IT tests, with correlations of -0.40 and -0.41, respectively. The auditory and Longstreth IT tests had the highest intercorrelation and were also the IT tests that correlated at relatively high levels with those psychometric tests that had lower discriminating power (APM and MHV) as can be seen in Table 4.4.

The IT-IQ correlation is likely to be reduced in this undergraduate population because of the attenuated range of ability. After correction for the attenuation of ability range (McNemar, 1955) some of the correlation values achieved levels of around -0.5 (Table 4.5). The AH5 test achieved some corrected correlations with inspection time scores at between -0.55 and -0.6.

#### **4.4 Discussion**

An AIT task was designed that was intended to offer some improvement over previous versions of AIT tasks. Analyses using oscilloscope tracings of waveforms established that the physical integrity of the items in the modified auditory inspection time test, especially those items of brief duration, was maintained when transferred from a specially constructed apparatus to a high quality reel-to-reel tape recorder. This allowed a standard test to be presented to all subjects. The performance curves of the subjects to the auditory inspection time test in this study appeared to approximate to the cumulative normal ogive predicted from the accumulator theory of visual discrimination and the results of Vickers, Nettelbeck and Willson (1972). However, no formal test of goodness-of-fit to such a model was undertaken, and the assessment of psychophysical test performance using

**Table 4.5**

Correlations between tests of psychometric intelligence and inspection time corrected for restriction of ability range.

	AH5 Verbal- Numerical	AH5 Non- Verbal	AH5 Total	Mill Hill Vocabulary	Advanced Progressive Matrices
Vertical Lines	-0.57	-0.52	-0.60	-0.04	-0.28
Horizontal Lines	-0.44	-0.59	-0.59	-0.36	-0.45
Longstreth	-0.13	-0.69	-0.54	-0.57	-0.53
Auditory	-0.22	-0.68	-0.57	-0.52	0.11

group data may be questioned (Levy, in press). There were no ceiling or floor effects for those subjects able to make the required pitch discrimination reliably. A significant minority of subjects appeared to be unable to make the discrimination required in the auditory inspection time test at any duration. This inability to discriminate between tones relatively widely separated in frequency in some student subjects was also found in students in other auditory tasks by Massaro (Massaro, 1976; Kallman and Massaro, 1979). Although pitch discrimination forms the basic discrimination required for the series of studies in this thesis, there are alternative modes of discrimination which can be used. There is a large literature on auditory backward masking, and alternatives such as loudness, duration or timbre might provide alternatives for those wishing to avoid the issue of individual differences in pitch discrimination (Kallman, Hirtle and Davidson, 1986).

In this study the IT tests were run as group tests. The visual IT tests were partially self-administered and, beyond familiarising the subjects with the stimuli and the computer and screen, there was minimal practice involved. These factors cut down on the time of testing and increased the convenience of the tests. Different psychophysical procedures were used to estimate the visual and auditory inspection times in this study. Both procedures were standard and there were unlikely to be discrepant results as a result (Levy, in press). The PEST procedure can reduce testing time considerably by 'homing in' on the stimulus duration where the subject's inspection time lies, but it results in subjects having variable numbers of trials per session (Stillman, 1989). The method of constant stimulus durations gives a more standard test, but subjects found this boring and, in contrast with the adaptive staircase, the form of the test fails to give the impression of periodically 'easing off'. The computer monitor screens, especially in the Longstreth test, could not provide a brief enough presentation to estimate subjects' inspection times below 20ms. Nevertheless, even with 40% of subjects receiving the same score on this test, significant correlations with cognitive ability tests were obtained.



Some of IT test intercorrelations were as high as the IQ test intercorrelations where, typically, the test-retest reliabilities are high, though the absolute level of these correlations was rarely above 0.5. Therefore, different visual IT tasks and the auditory and visual IT tasks share a modest amount of variance. It is noteworthy that the highest correlation between visual and auditory IT tasks was between AIT and the Longstreth task. The Longstreth version of IT does not involve the discrimination of two simultaneously-presented lines. Instead, subjects must decide which of two stimuli were presented before a mask. Therefore, it might be the case that there was a stronger temporal order aspect to the Longstreth task, which led to its having a higher correlation with AIT.

The auditory test correlated significantly with two of the three visual inspection time tests, but the magnitude of the three intercorrelations between visual and auditory tests was variable, from low to moderately high. The mean level of visual-auditory IT task intercorrelation approximated that found by Nettelbeck, Edwards and Vreugdenhil (1984), who reported a correlation of 0.39 between auditory and visual forms of the inspection time test in a study which included only non-handicapped, high ability adult subjects. A near-zero correlation between auditory and visual inspection time was reported by Irwin (1984). At this stage the best estimate of the correlation between visual and auditory IT in an adult population is probably in the region of 0.3 to 0.4. Although this is probably depressed by restriction of range in subject samples it would be unwise to estimate a 'true' value until larger scale studies have examined populations with normal distributions of ability. Therefore, the two tests appear to share some common variance. Whether the source of this variance is some form of general, cross-modal perceptual speed which has a neural basis or is a reflection of some subjects' sophistication in taking tests or in forming strategies to even such simple-seeming tests will be the subject of a later discussion. Nevertheless, the auditory-visual IT test intercorrelation matches the typical levels of correlation that are found when cognitive components from different tests are correlated with each other (Sternberg and Gardner, 1983).

Mackintosh (1986), Todman and Gibb (1985) and Vernon (1986) have suggested that finding a significant IT-IQ correlation might depend on the inclusion of retarded subjects which results in samples having an unrepresentatively large range of ability. Typically, the IT-IQ correlations in this experiment were lower than those reported by, for example, Brand and Deary (1982), and the levels rarely exceeded 0.3. In some cases the correlations were near to zero, although most were in the expected direction. When the IT-IQ correlations were corrected for restriction of ability range, some correlations approximated the -0.5 level which Nettelbeck (1987) and Kranzler and Jensen (1989) reckoned might be the 'true' IT-IQ correlation in a sample with a normal range of ability. However, corrected correlations must be treated with caution. Frearson, Barrett and Eysenck (1988) have reported that they tend to be over-estimates of true correlations. Therefore, they are not a substitute for testing in a large normal sample of the population. With the moderate to large sample sizes involved in the correlations reported in this study, many low correlations were significant. However, if a small sample size typical of many IT studies had been chosen, it can be calculated that one would fail to find correlations significant at the 5% level in approximately 70 out of every 100 experiments. Thus, while the presence of a significant correlation corroborates the IT-IQ relationship, the absence of one may not be used as evidence against it unless the sample size is large and the variance not restricted.

The few studies which have examined auditory and visual inspection times with intelligence in the same experiment were discussed in Chapter 2 (Brand and Deary, 1982; Irwin, 1984; Nettelbeck, Edwards and Vreugdenhil, 1986). Auditory IT correlated with ability tests at similar levels to the visual tests. With the inclusion of different ability tests and IT tests used in the present study it was hoped to be able to investigate the differences in IT-IQ correlations in verbal and non-verbal tests. Nettelbeck (1987) found that no specific type of intelligence had higher correlations with IT than others, and Nettelbeck,

Edwards and Vreugdenhil (1986) concluded that both visual and auditory IT were likely to be linked to general mental ability rather than to more specific types of ability. Including only those studies carried out using non-retarded adults Nettelbeck reported that, of 12 studies using verbal ability estimates, the average correlation with IT was -0.27 (7 were significant in the expected direction). Of 9 studies using performance ability estimates, the average correlation was -0.33 (7 were significant in the expected direction). Of 24 studies examining 'general' intelligence, the average correlation was -0.34 (16 were significant in the expected direction).

Cooper, Kline and MacLaurin-Jones (1986) investigated the relationships between visual IT and the Primary Mental Abilities and indicated that IT was more related to factors of visualisation and perceptual speed rather than to crystallised or fluid ability. That study, however, was based on results from only 20 undergraduates and the authors' conclusions were impressionistic and not based upon a factor analysis or on formal comparisons of correlations to check for significant differences. The results of the present study are in agreement with the findings of Nettelbeck's (1987) review. Both auditory and visual IT tests correlated significantly with both verbal and non-verbal ability scores. Auditory inspection time correlated significantly with the non-verbal section of the Alice Heim and the Mill Hill vocabulary scores, and at near-zero levels with verbal section of the Alice Heim and the Raven's APM. When the Longstreth IT test correlations were examined (Table 4.4) there were significant correlations with Mill Hill, Raven and AH5 (non-verbal) scores, but a near zero correlation with AH5 (verbal). The most parsimonious explanation of these results is that there is some general ability factor which shares variance with inspection time ability, and that there were no more non-significant correlations than might be expected given the power of the present study.

This study has indicated that the IT-IQ relationship is amenable to analysis under the conditions - group testing, restricted variance in IQ and high mean IQ - in which the

parameters of the tests will be most readily studied by experimental psychologists. It also established that the correlation between IQ and IT can be obtained despite the fact that in some tests many of these high IQ subjects were able to reach or exceed the minimum stimulus duration that the method of presentation could present.

The present study did not set out specifically to address the subject of strategy use in performing the IT tests, but the issue of 'specific' versus 'general' strategies used to explain the IT-cognitive ability correlation merits some mention. Certainly, the visual tests have been criticised for being vulnerable to strategies. One of the most frequently cited is the apparent motion strategy in which the mask overlaying the stimulus gives the impression of movement. This movement may be used by subjects to estimate the position of the long and short lines in the vertical lines task. Mackenzie and his co-workers (Mackenzie and Bingham, 1985; Mackenzie and Cumming, 1986; Mackenzie, Molloy, Martin, Lovegrove and McNicol, 1991) and Egan (1986, 1991) have discussed this form of strategy use which is a type of specific strategy. This is a potentially troublesome phenomenon and might indicate that some subjects find a 'short cut' to the solution of the discrimination problem. The existence of a single specific strategy is unlikely to account for the present results. It would be difficult to envisage a specific strategy that was useful in all three of the visual tests and in the auditory version of the test. This study, however, is not capable of answering a more general strategy hypothesis whereby the high IQ person is conceptualised as being generally more attentive, vigilant, motivated or generally more organised in the approach to the IT tests (Mackintosh, 1986; Ceci, 1990).

## Chapter Five

### Pitch discrimination, auditory inspection time and cognitive abilities

#### 5.1 Introduction

As discussed in Chapter 2, it has been suggested that AIT tasks might be indexing pitch discrimination ability rather than auditory information processing speed. At the very least, pitch discrimination ability might act as a confounding variable in AIT studies. The present study will address the question of whether speed of information processing in the auditory modality, as suggested by Brand and Deary (1982), or pitch discrimination ability, as suggested by Irwin (1984), is the main factor contributing to the auditory inspection time-cognitive ability correlation. Using the same AIT task as that used in Chapter 4, the following hypothesis will be tested here: AIT thresholds will still correlate with IQ scores with pitch discrimination scores obtained from standard tests of pitch discrimination are partialled out. In addition, a further attempt will be made to discover whether auditory inspection time has a closer relationship with verbal or with non-verbal cognitive ability scores.

The form of the auditory inspection time test devised for the present series of studies yielded AIT threshold estimate values which were longer than the very brief durations where the frequency content of the stimulus tones showed increasing similarity (Irwin, 1984). No undergraduate in the study reported in Chapter 4 had an auditory inspection time less than 30ms (mean 75.8ms, SD 27.5,  $n=80$ ). While the new AIT task has taken the AIT durations away from the region where pitch discrimination probably confounds duration as the key variable, the question of whether some AIT variance is attributable to individual differences in pitch discrimination remains to be addressed more directly. As discussed in Chapter 2, Irwin (1984) found that AITs correlated at about -0.5 with scores



on the Seashore test of pitch discrimination. But, it should be recalled that Irwin was using a white noise-masked AIT task, and that he did not report excluding those subjects unable reliably to perform the AIT task at long durations before analysing his data. The present study aims to discover whether individual differences in pitch discrimination ability continue to act as a confounding factor in the AIT-cognitive ability test association when these factors mentioned above are corrected.

An additional aim of the present study was to undertake a more formal analysis of the characteristics of the performance curves of subjects on the AIT task along three lines. First, the data represented in Figure 4.3 in Chapter 4 gave the impression that the pattern of subjects' responses on the AIT task was not incompatible with a cumulative normal ogive. However, this conclusion was based upon pooled subject data, and such decisions are more appropriately made on the basis of model-fitting to single subject data (Levy, in press). Second, Levy (in press) has suggested that the method of estimating IT thresholds that was used in Chapter 4 might be inefficient, and that the total number of errors on an IT task, or a threshold estimate derived from curve fitting procedures, might provide more efficient estimates of IT thresholds. Levy (in press) further indicated that IT-cognitive ability correlations might be underestimated when threshold estimates such as that used in Chapter 4 were calculated.

Third, two recent reports have found that chance responding in IT-type discrimination tasks persists beyond a stimulus time of 0ms, suggesting that some minimum level of stimulus duration must be presented to subjects before useful information may be extracted from a backward-masked stimulus. Levy (in press) has indicated that this holds for stimuli used in IT tasks, and Muise, LeBlanc, Lavoie and Arsenault (1991) reported this phenomenon with backward-masked alphanumeric stimuli. This is somewhat different from the theory of the accumulator model, which stated that above-chance level discriminations may be made at all non-zero stimulus times (Vickers, Nettelbeck and



Willson, 1972; Levy, in press).

In response to these issues, the present Chapter will: test individual subject data from the AIT test for goodness-of-fit to a particular model of discrimination performance; estimate AIT thresholds using three different methods and then observe their intercorrelation and their differential correlations with cognitive ability test scores; and examine where the AIT performance curve crosses the stimulus duration axis, i.e. estimate the highest stimulus duration that leads to chance levels of responding.

## **Study 1**

### **5.2 Method**

#### **5.2.1 Subjects**

Eighty-four second year psychology undergraduates were recruited for this study in order to take the auditory inspection time task. Fifty-nine of these subjects also provided cognitive ability test and pitch discrimination test data. All subjects were aged 19-23; 37 of the subjects who provided full datasets were female. No sex differences have been found in IT studies (Nettelbeck, 1987) and sex differences were not examined here. All subjects had normal or corrected-to-normal vision and reported no hearing deficits.

#### **5.2.2 Cognitive ability tests**

##### **Alice Heim 6 AG (Heim, Watts and Simmonds, 1983)**

This is an ability test which was constructed to discriminate among university level adults. It is a 40 minute group test intended for arts and general university students. It yields two subscores - verbal and numerical-diagrammatic - which may be added to give a total score.

The items are similar to those of the AH5 test used in Chapter 4. However, it was considered that the AH6 test offered a better separation of verbal and non-verbal abilities than the AH5 which combined verbal and numerical ability.

### **Seashore Pitch (Seashore, Lewis, Saetveit, 1956)**

This is a standard test of unspeeded pitch discrimination. It is a 50 item test where subjects were required to indicate the temporal order ('High-Low' or 'Low-High') of tone pairs which varied in pitch difference from 17 Hz to 2 Hz. The test began with easy items and progressed to the more difficult items. The tones lasted for a long duration (over 500ms) and were not backward- or forward-masked. The score derived from this test was the number of tone pairs discriminated correctly. The test was administered according to the instructions in the test manual, and as described below.

### **Auditory Inspection Time**

The test used here was the same as that described in Chapter 4. To recapitulate, it was composed of a fixed block of trials that progressed from long (easy) to short (difficult) stimulus durations. High tones were 880 Hz and low tones were 784 Hz. The auditory mask was a rapidly alternating series of 10ms bursts of both stimulus tones provided by a multivibrator circuit. Test stimuli were recorded on 1/4 inch tape and played on a UHER reel to reel tape recorder at high speed. Sound levels were 80 dB for stimuli and mask. Each item in the auditory inspection time task consisted of a cue tone (832 Hz) lasting 500ms, 1000ms of silence, a stimulus tone pair with no gap between them ('High-Low' or 'Low-High') and 1000ms of masking noise (described above). Thus there was no stimulus-mask gap and no interval between the two stimulus tones.

### **5.2.3 Procedure**

All subjects were tested on the Seashore pitch test one week prior to being tested on the

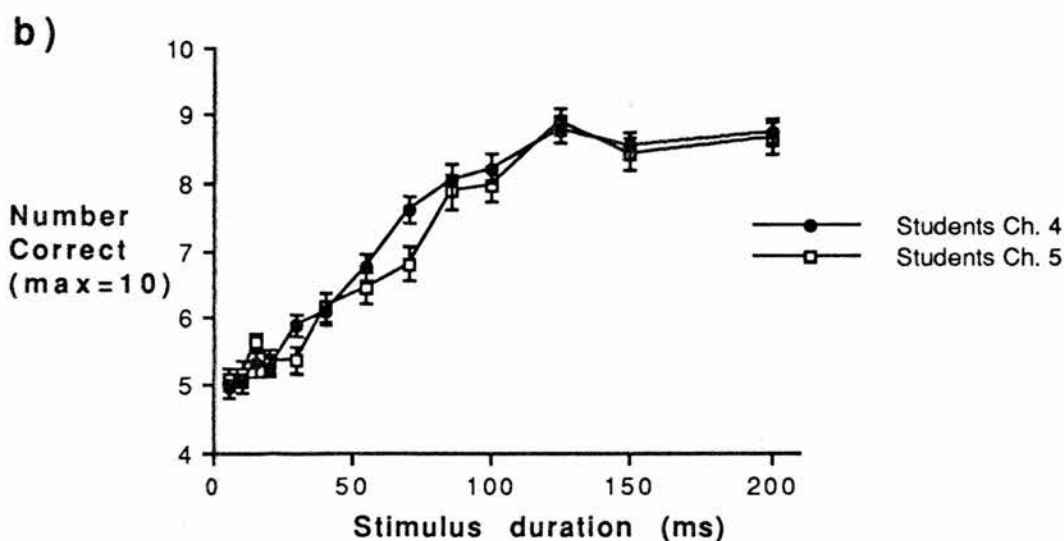
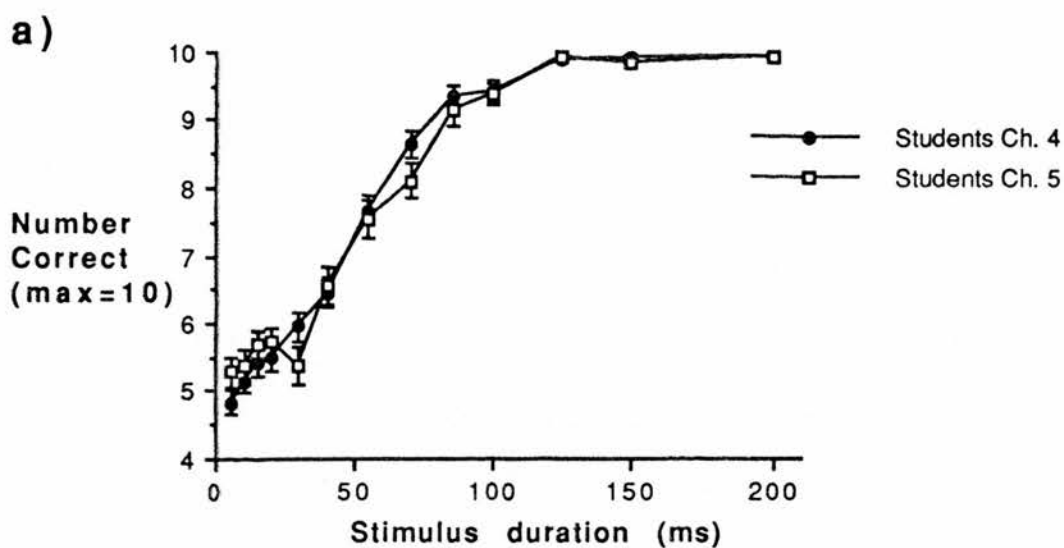
AH6 and the AIT tests. Seashore and inspection time testing was done individually in quiet basement cubicles and the stimuli were played through headsets which were pre-checked for sound level equality. AH6 was administered as a group test. Approximately half of the subjects took the AH6 prior to the AIT task and the others completed these in the reverse order. The AIT task was introduced in general terms to all subjects in a group fashion. Thereafter, detailed instructions and example items were supplied on a pre-recorded tape as before. Subjects were given 8 seconds between items to record their responses. Experimental trials came in 13 blocks of 10 stimulus pairs (each having 5 high-low and 5 low-high pairs) presented at progressively shorter durations, i.e. 200, 150, 125, 100, 85, 70, 55, 40, 30, 20, 15, 10, and 6ms. Those subjects who were unable to achieve 90% correct responses for the mean of the second practice and first experimental blocks were excluded from further auditory inspection time analysis.

### **5.3 Results**

#### **5.3.1 Auditory inspection time test performance**

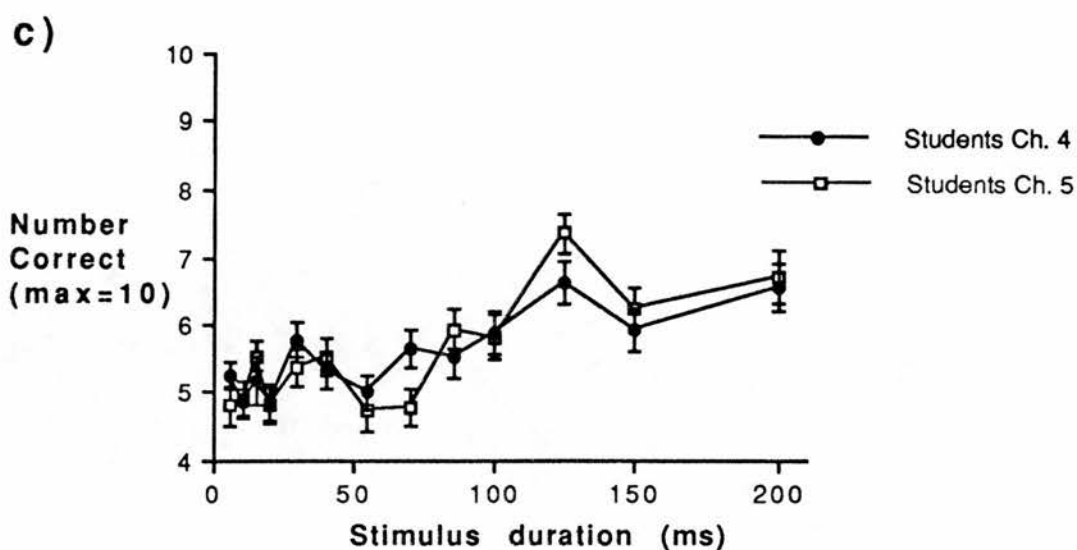
##### **Comparisons of data from the present study with those of Chapter 4**

Fifty-one out of 84 subjects (60.7%) were classified as being able to perform the discrimination in this study by comparison with 80 of the 120 subjects (66.7%) in the previous study. This difference between studies in the numbers of subjects defined as being able to perform the discrimination was not significant ( $\chi^2=0.53$ , d.f.=1, ns). Figure 5.1a shows the psychometric curves for the auditory inspection time task for those subjects who were defined as being able to perform the discrimination in the present study and in the study described in Chapter 4. The two curves were very similar, i.e. two independent groups of second year psychology students tested one year apart obtained very similar scores at the various stimulus durations. Figures 5.1b and 5.1c show the psychometric curves for all subjects in the two studies and for those subjects who were



**Figures 5.1a and 5.1b**

Mean (standard error) of AIT test items discriminated correctly at each of the 13 different stimulus durations for: a) those subjects classified as being able to perform the AIT test and, b) all subjects in the studies. Open square symbols represent the data from the student sample in this Chapter and the closed circle symbols represent the data from the student sample in Chapter 4.



**Figure 5.1c**

Mean (standard error) of AIT test items discriminated correctly at each of the 13 different stimulus durations for those subjects classified as not being able to perform the AIT test. Open square symbols represent the data from the student sample in this Chapter and the closed circle symbols represent the data from the student sample in Chapter 4.

classified as being unable to make the discrimination, respectively. Again, the curves were very similar, indicating reliability of the absolute mean scores at each duration across two groups who would be expected to perform at similar levels.

To check this similarity of the psychometric curves more formally, an analysis of variance was undertaken with Group as a between subjects factor (with two levels, i.e. whether subjects were tested for the study in Chapter 4 or Chapter 5) and stimulus duration as a within-subjects factor (with 13 levels, i.e. from 200 to 6ms). The null hypotheses under test were that there was no difference between the groups in the number of correct discriminations at each stimulus duration, and that duration made no difference to the number of correct answers on the AIT task. It was decided that the most appropriate analysis was to include the results for all subjects, rather than the results of only those subjects who were able to perform the task according to the criteria used in the study. The reasoning for this was as follows. There was a slightly greater proportion of subjects able to perform the task in the first year of testing, though this was not significant by chi square testing. This might come about for at least two reasons. First, it might be the case that some of the the people being classified as not being capable of making the discriminations had long inspection times, of about 200ms or above. Second, it might be the case that those classified as not being able to perform the discrimination were poorer at pitch discrimination. Therefore, the most conservative test of the above hypotheses was to include all subjects in both groups, lest the null hypotheses be supported because the present classification system led to an artefactual similarity of the two groups by excluding different numbers of subjects of both groups with long inspection times. Note that this potential exclusion of long inspection times would not lead to an artefactual increase in the correlations with IQ-type test scores, rather there might be slight deleterious effect that would lower correlations due to some restriction of the auditory inspection time threshold range.



The results of the ANOVA test are shown in Table 5.1. The effect of duration was highly significant ( $p < 0.001$ ), allowing the conclusion that longer stimulus durations led to significantly greater numbers of correct discriminations of tone order. The effect of group was not significant, confirming the impression that the two groups performed at similar levels over the two years. The group X duration interaction approached conventional levels of significance ( $p < 0.06$ ). Testing for simple effects showed that scores of the number of correct discriminations for only a single duration was significantly different between groups, i.e. at 70ms ( $F = 8.366$ , d.f. = 1, 1270,  $p = 0.004$ ). However, given the number of simple effects that might be compared, the chance of finding one such difference had a p value of about 0.052.

### **Comparing different AIT threshold estimates**

Because it was suggested in the introduction that the threshold measure used to estimate subjects' auditory inspection times in this Chapter and in the previous Chapter might be inefficient, and might underestimate the true AIT-cognitive ability test correlation, two additional measures were calculated in this study. First, the number of correct items in the block of AIT test trials was calculated, i.e. subjects were given a score out of 130 (the AIT test contained 10 trials at each of 13 stimulus durations).

Second, the data collected for each subject on the block of AIT trials was subjected to probit analysis (SPSS, 1990; Finney, 1971). In this analysis the proportions of correct responses at each stimulus duration are replaced with the value (the probit) of the standard normal curve below which the observed proportion of the area is found. The probit procedure then attempts to fit a straight line function to the relationship between the probit values and the log of the stimulus duration that resulted in the given probit value. The test of goodness-of-fit of the straight line function is given as a chi square value, with the degrees of freedom representing the number of stimulus durations minus the number of estimated parameters. If the chi square value is non-significant then this suggests that the

**Table 5.1**

ANOVA testing of undergraduate samples from Chapter 4 (n=117) and this Chapter (n=84) on number of correct responses on the auditory inspection time task. Groups (with two levels) was a between subjects factor and duration (of stimulus) was a within subjects factor (with 13 levels).

Source of variation	d.f.	Sum of squares	Mean square	F	p
Group	1	7.894	7.894	0.446	.50
Error	199	3521.951	17.698		
Duration	12	5010.197	417.516	151.316	0.000
Group x Duration	12	55.927	4.661	1.689	0.063
Error	2388	6589.065	2.759		

**Table 5.2**

Alice Heim 6 (AH6) and pitch discrimination scores (Mean + SD) for those undergraduates who could and could not perform the auditory inspection time task.

	Could do auditory inspection time (n=34)	Couldn't do auditory inspection time (n=25)	t	p
AH6 Verbal	15.8 (4.3)	15.8 (4.9)	0.01	ns
AH6 N+D	14.2 (4.2)	13.8 (4.7)	0.32	ns
AH6 Total	30.0 (8.0)	29.6 (8.2)	0.18	ns
Seashore pitch	44.3 (2.8)	36.7 (6.0)	5.94	<0.001
Auditory Inspection Time (ms)	77.9 (26.1)			

model fits well. If the chi square value is significant this might indicate that the relationship was non-linear or that the spread of the data points about the regression line was heterogeneous. Regression coefficients and their standard errors are obtained from the analysis. The coefficient divided by the standard error may be used as a z score to indicate whether the regression coefficient is significant (values greater than 1.96 were used to indicate significant coefficients, at  $p < 0.05$ ) (Bentler, 1989). Probit analysis also gives the expected proportion of correct items for any given duration, with confidence limits. In this study the 90% threshold, i.e. the stimulus duration at which 90% correct responses were expected, derived from the probit analysis was compared with the other two estimates of AIT performance. The stimulus duration at which 50% of responses were expected to be correct was of interest because it might indicate the minimum stimulus duration at which useful information may be extracted from the stimulus.

The probit analyses results for the 84 subjects who provided full AIT data in the study are presented in Appendix 3. According to criteria presented earlier, 51 subjects were classified as being able to perform the AIT task and 33 were unable to perform the test. Appendix 3 lists: subjects' identity codes; probit-derived regression coefficients and their standard errors; the value obtained when each coefficient was divided by its standard error; the chi square for each subject's model with its significance level; the 90% and 50% thresholds for the probit model; and the 95% confidence intervals for the 50% threshold. The chi square values were non-significant for 34 of the 51 subjects who could do the AIT task, indicating that the data from 66.7% of these subjects fitted the model well. As expected, regression coefficients were significant for all of the subjects, indicating that all of these subjects' scores benefitted from increases in stimulus duration. The lack of confidence intervals for the 50% threshold for four of these 51 subjects indicated that their data were heterogeneous. For the 47 subjects whose data were not heterogeneous, the mean duration corresponding to the 50% threshold was 12.18ms (SD 3.9) and for only a single subject did the 95% confidence limits on the 50% thresholds include the value of

0ms.

On the other hand, only seven of the 33 subjects (21.1%) unable to do the AIT task had data that were not heterogeneous (Appendix 3). In only eight of these latter subjects (24.2%) was the regression coefficient significant, indicating that only a small proportion of these subjects had scores on the AIT task which benefitted from increases in stimulus duration. Almost all of the subjects who could not perform the AIT task had non-significant chi square values, and this appeared to be because the data for most subjects conformed to a horizontal line, with the different stimulus durations resulting in largely chance levels of responding.

The relationships among the three estimates of AIT performance were examined in those subjects who were able to perform the AIT task and whose data were not heterogeneous according to probit analysis ( $n=47$ ). A logarithmic transformation of the probit 90% thresholds was used to obtain a near to normal distribution of AIT threshold scores. AIT thresholds using the criteria described in the previous Chapter correlated at 0.87 with probit-derived 90% thresholds and at -0.79 with the number of correct items (both  $p<0.001$ ). Probit-derived 90% thresholds correlated at -0.90 ( $p<0.001$ ) with the total number of correct items.

### **5.3.2 Correlations among auditory tests and cognitive ability tests**

Table 5.2 shows the summary data and t test results for those subjects who could ( $n=34$ ) and those who could not ( $n=25$ ) perform the auditory inspection time test from the 59 subjects who were tested on the the cognitive ability tests and the pitch discrimination test. The basis for this separation was the score on the masked and unmasked items where the duration of the tones was 200ms. There was no difference in AH6 scores or subscores, but the Seashore pitch discrimination scores differed significantly between the two groups ( $p<0.001$ ).

A correlation analysis was performed on the test variables for those subjects who could perform the auditory inspection time test. Table 5.3 shows the high correlation between the AH6 subtests. The correlation between pitch discrimination and AH6 was near to zero and non significant. The correlation between AIT and AH6 total score was -0.39 ( $p < 0.05$ ), with AH6 verbal score was -0.45 ( $p < 0.01$ ) and with AH6 numerical and diagrammatic score was -0.27 ( $p$  approx. 0.1). The correlation between auditory inspection time and pitch discrimination was in the direction which indicated that those subjects who were better on the inspection time test also obtained better Seashore pitch scores, but the association was non-significant ( $r = -0.20$ ). With pitch discrimination scores held constant, the partial correlation between auditory inspection time and AH6 total score was -0.38, i.e. it remained similar to the raw correlation.

Table 5.4 shows the correlations between AIT ability thresholds derived using the three different estimating procedures discussed above and the Alice Heim scores. The number of subjects was greater than that reported in Table 5.3 because Seashore data was not available for some subjects, who were not included in the analysis in Table 5.3. It can be seen in Table 5.4 that the highest correlations with cognitive ability occurred with the logarithmic transformations of the probit-derived 90% threshold estimates. These correlations exceeded those involving the AIT thresholds used above and in Chapter 4 by about 0.02. Using a threshold based upon the total number of correct discriminations in the AIT test resulted in levels of correlations that were slightly lower than those obtained from the other two AIT estimates.

### **5.3.3 Reliability of the AIT test scores**

AIT test response sheets for the complete subject sample were re-scored, and separate scores for the odd-numbered and even-numbered items were obtained. The correlation for the two halves was 0.795, which gives a split-half reliability estimate of 0.886 for the AIT

**Table 5.3**

Correlations among Alice Heim 6 (AH6), auditory inspection time and Seashore pitch test scores for subjects able to perform the auditory inspection time task (n=34).

	AH6 Verbal	AH6 N+D	AH6 Total	Seashore pitch test	Auditory Inspection Time
AH6 Verbal	-	0.77***	0.94***	-0.06	-0.45**
AH6 N+D		-	0.94***	-0.09	-0.27
AH6 Total			-	-0.08	-0.39*
Seashore pitch test				-	-0.20

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

**Table 5.4**

Correlations between three different estimates of AIT thresholds from data in used in the present study and Alice Heim 6 cognitive ability scores (n=46).

	Alice Heim Verbal	Alice Heim Numerical & Diagrammatic	Alice Heim Total
AIT(ms) as computed in Chapter 4	-0.348*	-0.215	-0.304+
Number of correct AIT test items	0.316+	0.229	0.295+
Logarithmic transformation of probit-derived 90% threshold measures	-0.370**	-0.240	-0.330*

+  $p < 0.1$ ; \*  $p < 0.05$ ; \*\*  $p$  approx 0.01



test total score.

## 5.4 Discussion

Two independent, moderately large student groups were tested on the same auditory inspection time test one year apart. The selection criteria for students in the psychology curriculum had not changed during that time. Therefore, given that the conditions of testing were replicated as closely as possible, the same proportion of subjects should be able to perform the test and the number of correct discriminations at each duration should be similar across the years. Students in the two years were tested in the same quiet basement cubicles, using the same sound distribution system, stimulus tape, testers and the same headphones which had sound levels checked with the same sound meter. The proportion of subjects classified as being able to perform the discrimination in the two studies did not differ significantly, and there were no overall between-group differences on the number of correct discriminations of tone order in the auditory inspection time test. Simple effects testing indicated that, in the second year of testing, the number of correct order discriminations in the task was lower at a stimulus duration of 70ms, but one significant simple effect was about what might be expected due to chance.

Three methods of estimating AIT ability from the same data were correlated very highly. The probit analyses provided corroboration for the assumption that most subjects' responses in the task conformed to a normal ogive, although competing models would have to be tested to give more powerful support to this claim (Levy, in press). Almost all subjects who could perform the AIT task were scoring at chance levels at stimulus durations which were significantly greater than 0ms, indicating that a minimum stimulus duration was required to afford information about the stimulus difference. This conforms to the two-stage model of information intake proposed by Muise, LeBlanc, Lavoie and Arsenault (1991). This model states that there is a minimum stimulus duration,  $T_{lag}$ , below which no useful information may be extracted from a stimulus. Thereafter, the

model states that the increase in information extracted from increases in stimulus duration may be described by a negatively accelerated exponential function. To test this model versus others would require very large numbers of trials obtained from single subjects, and must remain a task for future research.

Therefore, the present task appears to offer stable results in similar groups across time, and the data generally appear to fit a model that was developed to account for visual inspection time discrimination. In addition, the split-half reliability estimate for the total score on the AIT test was high, having a split-half reliability of 0.886. There is some evidence from the performance curves to support the existence of a minimum stimulus duration, of about 13ms, below which no information may be extracted to aid discrimination processes.

This study attempted to examine the effect on the correlation between scores on the auditory inspection time test and an IQ-type test scores when differences in pitch discrimination ability were controlled by partial correlation. The correlation between total score on the AH6 test and auditory inspection time was -0.39, which is very similar to the correlation of -0.38 reported between auditory inspection time and Raven's Advanced Progressive Matrices (Nettelbeck, Edwards and Vreugdenhil, 1986). With respect to the contribution made by pitch discrimination ability to the auditory inspection time-mental ability test correlation, pitch discrimination *per se* did not appear to be the basis of the correlation, because the raw correlations between cognitive ability scores and Seashore pitch scores were not significantly different from zero. Moreover, the Seashore-AIT correlation was non significant, and the AIT-AH6 correlation decreased very little when pitch discrimination ability differences were partialled out. It was also expected that pitch discrimination ability might play a relatively small part in contributing to the between subject variance in the AIT task used here because AIT thresholds were out of the range within which other researchers have indicated that frequency similarity of the stimulus tones is a potential problem (Irwin, 1984). These considerations appear to indicate that

the auditory inspection time test is not indexing pitch discrimination primarily. This provides some discriminant validity for the claim that the AIT test is indexing speed of auditory processing.

Although the present study did not find a significant association between pitch discrimination and cognitive ability, there are others which have done so. Some of the early studies were reviewed in Chapter 3. In addition to these, some more recent studies have replicated the slight association between pitch discrimination ability and psychometric intelligence. McLeish (1950) carried out a factorial study of the different musical tests included in the Seashore battery on 100 psychology students, and found that the average correlation between the six Seashore tests and Cattell IQ scores was 0.17. The correlation with Seashore pitch test was 0.22. Lynn, Wilson and Gault (1989) found that unspeeded pitch discrimination as measured by the Bentley (1966) test had a loading of 0.49 on the first unrotated component of a principal components analysis of several cognitive and musical ability tests in 217 10-year-olds. Because the pitch test was unspeeded in nature, Lynn, Wilson and Gault (1989) concluded that it was accuracy rather than speed of information processing through sensory tracts that was an important aspect of "the neurophysiology of g".

Karlin (1942) carried out a large scale factorial study of auditory abilities on 200 high school students. Pitch discrimination test scores from the Seashore battery correlated at only 0.147 with IQ estimates based on the Otis or Henmon-Nelson tests. However, Karlin (1942) reckoned that it was also important to discover the shortest stimulus durations at which subjects could make accurate pitch discriminations,

In the short-impulse pitch discrimination test it was considered of interest to discover what relationship there might exist between the duration threshold necessary for accurate pitch judgements and other forms of pitch judgments above the duration threshold.

Karlin's (1942) short impulse pitch discrimination test varied the duration of two tones of constant intensity and complexity, and of "supra-liminal" pitch difference, and determined the duration required by subjects in order to make easy pitch discriminations. This was arguably the earliest attempt to measure a type of auditory inspection time in the literature, though few details of the test were given. Scores on the short impulse test correlated at 0.213 with IQ and at 0.628 with Seashore pitch discrimination test scores in the high school students. Therefore, speeded auditory processing had a higher correlation with IQ than did unspeeded discrimination, though both correlations were low, and the speeded and unspeeded discrimination tests had moderately high intercorrelations.

It is noteworthy that 42.4% of the subjects were unable to make the pitch discrimination reliably enough for their data to be included in the AIT analysis. This is in agreement with the previous Chapter's results. The present study did not extend the duration of the tones in the auditory inspection time test beyond 200ms, and it may not be stated definitively that some of these subjects did not have inspection times at longer durations. Nevertheless, they had similar AH6 scores to those who could do the auditory inspection time task, but had significantly lower pitch discrimination scores. It appears to be prudent to suggest that, when it is considered to be important to control for pitch discrimination ability in a subject sample, pitch discrimination screening of subjects participating in auditory inspection time studies should take place. This might be analogous to the practice of ensuring that subjects in visual inspection time studies are screened for their ability to make adequate visual discriminations. Once subjects had attained a certain level of pitch discrimination, it did not appear to play a significant part in auditory inspection time performance, i.e. it might have acted principally as a threshold factor on, rather than as a correlate of, auditory inspection time performance.

In these undergraduates pitch discrimination ability was not related to AH6 scores. By eliminating those subjects who were unable to perform the auditory inspection time task at

any duration included in the test, and whose main problem was assumed to be one of poor pitch discrimination ability, the apparent problem of unduly large auditory inspection time standard deviations as reported by Deary (1980), Irwin (1984) and Nettelbeck, Edwards and Vreugdenhil (1986) appears to have been overcome. After partialling out pitch discrimination test scores from the auditory inspection time-mental ability correlations, little difference in their magnitude was noted. Therefore, it may be hypothesised tentatively that temporal resolution might be the key aspect of auditory inspection time that contributed to the correlation with cognitive ability test scores. However, to examine further the possibility that any AIT-IQ correlation in *children* is due to pitch discrimination differences, as Irwin's (1984) findings indicated, a similar study to the one described above was carried out on a large group of children.

## Study 2

### 5.5 Subjects

Sixty boys (mean age 11 years 4 months, SD 4.2 months) and 59 girls (mean age 11 years 6 months, SD 5.1 months) recruited from the primary seven classes of an Edinburgh primary school took part in this study. In response to letters taken home by all pupils, no parents refused to give permission for their children to take part. Therefore, the children comprised an unselected sample. All subjects had normal or corrected-to-normal vision and none of the parents of the children or the children themselves reported their having hearing difficulties, though this was not tested formally.

### 5.6 Method

#### 5.6.1 Cognitive Ability Tests

Two verbal ability tests, one non-verbal intelligence test and one mathematical reasoning

test were administered to each subject.

### **Verbal Reasoning Test 89**

The Verbal Reasoning Test 89 from the Moray House series was used (Godfrey Thomson Unit, 1970). It consists of 100 items involving series, classification and logical problems, most of which involve words but some of the series items involve number concepts. This yielded an age-corrected verbal reasoning quotient (VRQ).

### **Mill Hill Vocabulary**

The Mill Hill Vocabulary Test - Form 1 Junior, Parts A and B (i.e. definitions and synonyms; Raven, Raven and Court, 1982) involved subjects both in defining words and in selecting the closest synonym for words from a number of alternatives. This test yielded an age-corrected verbal IQ (Mill Hill IQ).

### **Raven's Progressive Matrices**

The Raven's Progressive Matrices (Raven, 1958) test was administered as a 40 minute timed test, and this was age-corrected to give a non-verbal IQ (Raven IQ). This involved subjects in selecting, from a number of alternatives, a shape that would complete a pattern which had a logical basis. This is a 60-item test and the items are divided into five sections. The difficulty of the items increases within and between sections.

### **Mathematical Reasoning**

The Mathematics 4a Test from the Moray House series was given (Godfrey Thomson Unit, 1969), and it was age-corrected to give a mathematical reasoning quotient (MRQ). This test includes mental arithmetic and mathematical problems.

### **5.6.2 Pitch Discrimination Test**

The pitch discrimination subtest of the Bentley Tests of Musical Ability (Bentley, 1963)



was used. The pitch discrimination test is a twenty item test involving tone pairs which are played consecutively, with no backward mask. Stimulus duration of the tones is about 0.5 second, there is a silent gap of a similar magnitude between the two tones, and the subject is required to state whether the second tone is higher, lower or the same as the first tone. The test begins with tone pairs which are widely separated in pitch and items become progressively more difficult as the test proceeds.

### **5.6.3 Auditory Inspection Time**

The schoolchildren undertook the same test as did the undergraduates in the study reported above. The group instructions prior to testing were more extensive and checks were made upon individuals in order to ensure that subjects understood the nature of the test. Otherwise, the taped instructions and the test setting were identical to that undertaken by the student samples.

### **5.6.4 Procedure**

The Bentley pitch discrimination test was administered to the subjects in groups of 20 to 25 in quiet classrooms in their primary school, according to the instructions in the testing manual. No formal sound levels were taken for the Bentley pitch test administration, though the administrator made checks to ensure that all testees could hear the stimuli clearly and that they understood the test principles. The mathematical and verbal reasoning tests from the Moray House series were administered in classrooms as group tests to similarly-sized groups. These tests were administered under examination conditions by Mr B. Head, the Headmaster of the Upper Primary School. Testing at the school was performed approximately two months prior to subjects being tested on the IQ and auditory tests detailed below.

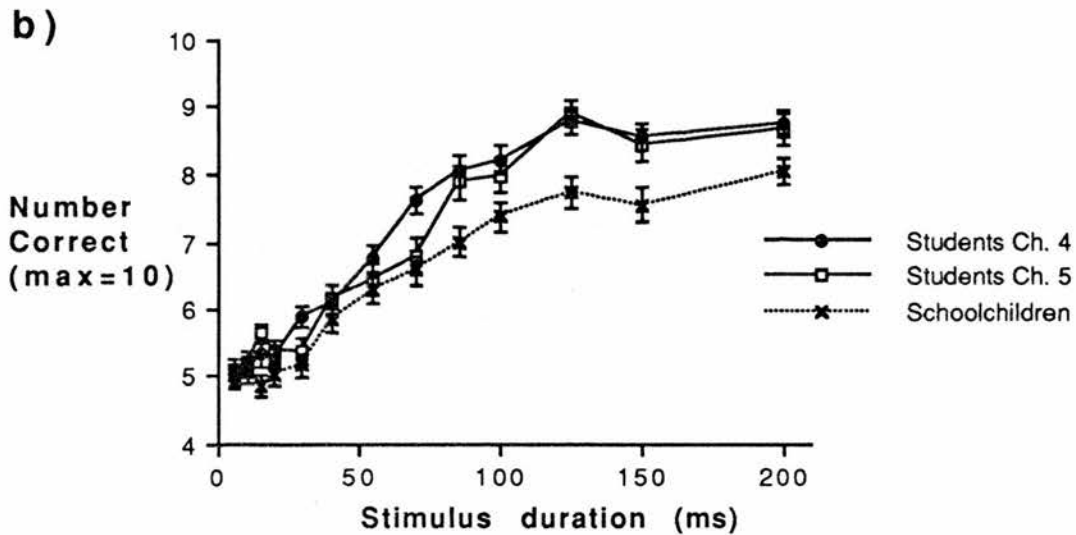
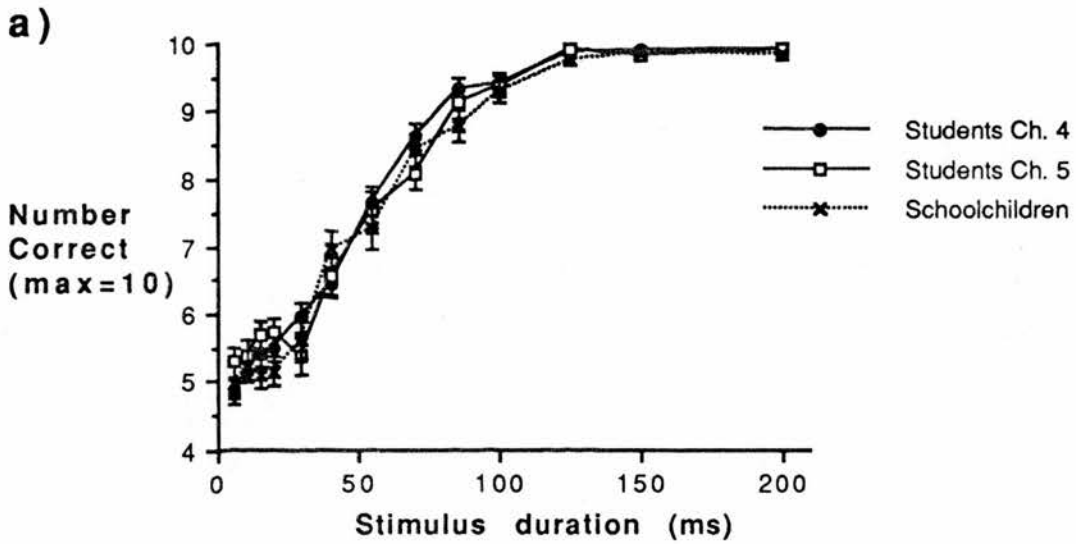
Subjects visited the Department of Psychology in groups of about 10 in order to undertake the remainder of the tests. Testing took half a day to complete. Raven, Mill Hill and

auditory inspection time tests were administered in the same situations as reported in the previous experiment. All subjects undertook the tests in the same order: Raven, Mill Hill and, after a 30 minute break for refreshments, a second administration of the Bentley Pitch Test followed by the auditory inspection time test. All ability test raw scores were converted to IQ-type scores and the auditory inspection time was scored as described in Chapter 4.

## **5.7 Results**

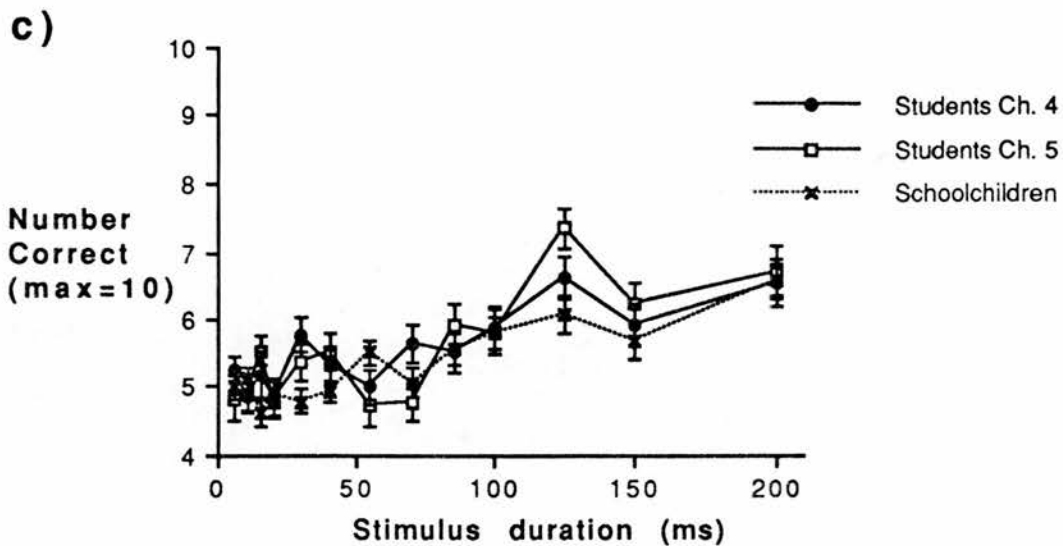
### **5.7.1 Auditory inspection time test responses**

Twenty nine of the 60 boys and 24 of the 59 girls were able to perform the auditory inspection time task according to the criteria applied to the student samples in this and the previous Chapter. This proportion was not significantly different from the 34 out of 59 students who could perform the task (chi square = 2.21,  $df=1$ , ns). Figures 5.2a to 5.2c show the group performance curves for the schoolchildren on the AIT task. Data from the schoolchildren were superimposed on the data from the undergraduate samples which provided the data for the studies in Chapter 4 and the present Chapter. Figures 5.2a and 5.2c show, respectively, that those children who were classified as being able and not able to perform the AIT task had similar psychometric curves to those found in the student samples. However, when the mean performance scores of all of the subjects in the present study were compared with the complete student samples, the children appeared to have lower numbers of correct responses, especially at stimulus durations of 85ms and longer. The data from all schoolchildren was compared to that of the student AIT data samples in the present study and in Chapter 4 using a two way ANOVA. Groups was a between subjects factor (with three levels, i.e. the two student groups and the schoolchildren) and duration as a within subjects factor (with 13 levels, i.e. the 13 stimulus durations). The results of this analysis are summarised in Table 5.5 which shows that the effects of group and duration and their interaction were all very highly significant. Post-hoc testing showed



**Figures 5.2a and 5.2b**

Mean (standard error) of AIT test items discriminated correctly at each of the 13 different stimulus durations for: a) those subjects classified as being able to perform the AIT test and, b) all subjects in the studies. Open square symbols represent the data from the student sample in this Chapter and the closed circle symbols represent the data from the student sample in Chapter 4. The cross symbols with the dashed lines represent the data from schoolchildren in the present study.



**Figure 5.2c**

Means (standard error) of AIT test items discriminated correctly at each of the 13 different stimulus durations for those subjects classified as being not able to perform the AIT test. Open square symbols represent the data from the student sample in this Chapter and the closed circle symbols represent the data from the student sample in Chapter 4. Cross symbols with dashed lines represent the data from the schoolchildren in the present Chapter.

**Table 5.5**

ANOVA testing of the sample of schoolchildren used in the present Chapter (n=119) and undergraduate samples from Chapter 3 (n=117) and this Chapter (n=84) on number of correct responses on the auditory inspection time task. Groups (with three levels) was a between subjects effect and duration (of stimulus) was a within subjects effect (with 13 levels).

Source of variation	d.f.	Sum of squares	Mean square	F	p
Group	2	271.130	135.565	7.315	.0008
Error	317	5874.995	18.533		
Duration	12	6933.001	577.750	201.020	0.000
Group x Duration	24	166.739	6.947	2.417	0.0001
Error	3804	10933.045	2.874		

**Table 5.6**

Mental ability test and Bentley pitch test scores and ages (Mean + SD) for those subjects in the sample of schoolchildren who could and could not perform the auditory inspection time task.

	Could do auditory inspection time (n=53)	Couldn't do auditory inspection time (n=66)	t	p
Raven IQ	116.0 (10.7)	111.6 (10.2)	2.31	<0.05
Mill Hill IQ	118.7 (10.4)	113.3 (12.4)	2.50	=.01
Verbal Reasoning Quotient	116.1 (12.4)	108.1 (10.4)	3.81	<.001
Mathematical Reasoning Quotient	111.9 (11.2)	106.2 (10.6)	2.81	<0.01
Bentley Pitch test 1	16.9 (1.8)	14.3 (2.7)	6.11	<0.001
Bentley Pitch test 2	17.1 (2.6)	14.9 (2.8)	4.45	<0.001
Age (months)	140.8 (3.5)	140.9 (5.5)	0.06	ns
Auditory Inspection Time (ms)	75.7 (31.2)	-		

a significant difference between the sample of schoolchildren and the two student samples which was beyond the  $p<0.01$  level. Simple effects were tested and there were significant differences at 200ms ( $p=0.02$ ), 150ms ( $p<0.001$ ), 125ms ( $p<0.001$ ), 100ms ( $p=0.008$ ), 85ms ( $p<0.001$ ), 70ms ( $p<0.001$ ), 30ms ( $p=0.028$ ) and 15ms ( $p=0.019$ ).

### **5.7.2 Group comparisons**

Girls scored significantly better on the pitch discrimination task at the second sitting ( $p<0.05$ ) and had higher mathematical reasoning quotients ( $p<0.01$ ) than the boys, otherwise the two groups were very similar and data analysis was performed on the whole group. Table 5.6 has the summary data for schoolchildren divided into those who could and those who could not perform the auditory inspection time task. This separation was performed using the same criterion as that used in Chapter 4 and in the first study in this Chapter. There was no age difference between the two groups. Pitch discrimination was significantly different between the groups ( $p<0.001$ ). Unlike the student group, both the verbal and non-verbal IQ test scores and the verbal and mathematical reasoning scores were significantly different in the two groups with those able to perform the auditory inspection time task having higher mean scores.

Table 5.7 shows the Pearson correlations among the cognitive ability, Bentley Pitch and AIT tests. All ability test intercorrelations were positive and significant. Test-retest reliability on the Bentley Pitch test was 0.52 ( $p<0.001$ ). There were 8 correlations between pitch discrimination ability and mental ability (pitch discrimination estimated at two sittings versus four ability tests) with a range of 0.04 to 0.29 (mean=0.15). When corrections were made to these correlations to take account of the moderate test-retest reliability of the Bentley Pitch test, the range for the corrected correlations was 0.06 to 0.45 (mean=0.25). There appeared to be a difference between the verbal and non-verbal ability test correlations with pitch discrimination (Table 5.7). Raven and MRQ correlations with pitch discrimination ability were all near to zero and non-significant. Three of the



**Table 5.7**

Correlations among mental ability test, Bentley pitch test and auditory inspection time scores for schoolchildren sample (n=119, except for correlations involving auditory inspection time where n=53).

	Raven IQ	Mill Hill IQ	VRQ	MRQ	Bentley Pitch 1	Bentley Pitch 2	Auditory Inspection Time
Raven IQ	-	0.43***	0.66***	0.64***	0.10	0.09	-0.26*
Mill Hill IQ		-	0.55***	0.41***	0.29**	0.22*	-0.36**
Verbal Reasoning Quotient (VRQ)			-	0.82***	0.14	0.21*	-0.28*
Mathematical Reasoning Quotient (MRQ)				-	0.04	0.12	-0.24+
Bentley Pitch test 1					-	0.52***	-0.26*
Bentley Pitch test 2						-	-0.11

+p <0.1; \* p<.05; \*\* p<.01; \*\*\* p<.001

**Table 5.8**

Partial correlations between ability test scores and auditory inspection time controlling for pitch discrimination in schoolchildren (n=53).

	Raven IQ	Mill Hill IQ	VRQ	MRQ
Controlling for Bentley Pitch test 1	-0.25+	-0.32*	-0.27*	-0.28*
Controlling for Bentley Pitch test 2	-0.26+	-0.35**	-0.28*	-0.25+

+ p<0.1; \* p<0.05; \*\* p<0.01

four verbal ability test-pitch discrimination correlations were significant and in the expected direction. Therefore, superior pitch discrimination appeared to convey some advantage on tests of vocabulary and verbal reasoning that was not apparent in the undergraduate sample.

Pitch discrimination scores correlated significantly with AIT test thresholds at the first pitch test and in the same direction, but non-significantly, at the second testing of pitch. All correlations between mental ability test scores and AIT test thresholds were negative and significant (Table 5.7). As was found in the undergraduate sample in the present study, it was the verbal IQ which correlated at a higher level with thresholds from the auditory inspection time test ( $-0.36$ ,  $p < 0.01$ ). Table 5.7 shows that the school performance-related tests of mathematical and verbal reasoning correlated significantly with AIT test thresholds. Table 5.8 shows the partial correlations between ability test scores and auditory inspection time thresholds when the potentially confounding effect of pitch discrimination ability was partialled out. The correlations altered little from the original results, with the Mill Hill-AIT correlations becoming  $-0.32$  and  $-0.35$  and the Raven IQ-AIT correlations becoming  $-0.25$  and  $-0.26$ , depending, respectively, on whether the first or second pitch discrimination test scores were partialled out.

## **5.8 Discussion**

The psychometric curves for AIT responses of groups of 11-year-old schoolchildren able and unable to perform the AIT task were very similar to the curves for undergraduates, but the group as a whole made fewer correct discriminations of temporal order. The present study was not designed to test age differences in AIT, and the meaning of this difference is not clear, because the students were both older and almost certainly had higher mean IQ scores than the children. Therefore the effects of age and ability were probably confounded. Nettelbeck and Wilson (1985) demonstrated that, whereas 11-year-olds had significantly briefer visual inspection times than 8-years-olds, the 11-year-olds did not

differ from adults. This was in agreement with the review of the developmental backward masking literature by Nettelbeck and Brewer (1981) who concluded that inspection time probably increases with age until about 10 years, and with the view of Anderson (1988) who found that visual inspection times do not improve with increase in age through childhood. However, Nettelbeck and Wilson (1985, study 1) used only ten 11-year-olds and 10 adults. As in the present study, the children appeared to be unselected for cognitive ability and the adults were university students. Therefore, neither the first study of Nettelbeck and Wilson (1985) nor the present study may be used as a definitive test of whether IT improves after 11 years.

Study 3 by Nettelbeck and Wilson (1985) tested seven groups of ten children, aged 7 years and 4 months to 13 years and 2 months and a group of 10 university undergraduates on visual inspection time. They concluded that,

...these data suggest a marked decrease in inspection time up to around age 13 years, with the possibility of less marked change beyond.

This would be in agreement with a conclusion, albeit with the caveats raised above, from the present study that AIT improves from age 11 to adulthood. However, this is somewhat undermined by the fact that Table 4 of Nettelbeck and Wilson (1985) shows that the 11-year-olds had mean ITs and IQs of 132ms and 107, respectively, whereas the corresponding figures for the 13-year-olds were 162ms and 123. Even if the differences between the schoolchildren and student groups in the present study were due to age differences in AIT and not due primarily to differences in mean cognitive ability level, there are other considerations that might prevent a straightforward conclusion to the effect that AIT improves from 11 years to young adulthood. Ross and Ward (1978) warned that age differences that are found in visual backward masking tests might be due to differences in learning or attention to the task.

The present study attempted to investigate the relationship between auditory inspection time and mental abilities in children in a fashion that might improve upon some of the aspects of Irwin's (1984) study. By using a modified auditory inspection time test, and as a result of screening for inability to perform the auditory inspection time task at 200ms, even when the stimuli were unmasked, the problems of very long auditory inspection time durations and high auditory inspection time standard deviations appear to have been reduced. In other studies, these factors might have introduced a marked contamination of temporal resolution speed with pitch discrimination ability. Pitch discrimination had uncorrected correlations of about 0.15 with IQ scores in this sample of schoolchildren. Partial correlations between IQ and auditory inspection time, controlling for pitch discrimination, differed minimally from the original correlations. It appeared that, whereas auditory inspection time did correlate at low levels with pitch discrimination ability, the auditory inspection time-IQ correlation was unlikely to be caused principally by the ability to make fine pitch discriminations. Therefore, the data from this sample of children, at least, indicate that the abilities of sensory discrimination and temporal resolution appeared to be independent correlates of psychometric ability tests.

It was important to examine the data for evidence which might indicate that particular abilities were related to IT. The results reported in Chapter 4 did not appear to indicate that any one type of ability showed higher correlations with auditory inspection time scores than the others. The data from experiment two in this Chapter indicate that the high-verbal child has an advantage on certain sensory indices - he or she has somewhat better pitch discrimination ability and a shorter AIT threshold. The child scoring highly on non-verbal or mathematical tests appeared to have no particular advantage in pitch discrimination and to have less of an advantage overall in the auditory inspection time task. This concurs with the finding of Nettelbeck and Young (1990) that, "a higher correlation between IT and VIQ [Wechsler Intelligence Scale for Children-Revised Verbal IQ] is probably more characteristic of children". It also concurs with the extensive research of Tallal (1989) who

has indicated that the ability to make accurate discriminations of temporal order in both the auditory and the visual modalities is related to superior language development.

### **5.9 General discussion of studies 1 and 2**

The two studies in this Chapter corroborate those few existing reports which indicate that auditory inspection time has a significant association with cognitive ability. Significant correlations were obtained despite the fact that the students and the schoolchildren were of above average ability. In addition, the student sample was restricted in ability range.

Auditory inspection time measurements in previous research might have confounded pitch discrimination ability with auditory information processing speed. By using a more effective mask and by eliminating any inter-tone gap, which might have allowed rehearsal of the first stimulus tone while awaiting the arrival of the second, the *lowest* obtained auditory inspection times (85% threshold estimates) increased to about 30ms without altering the proportion of subjects able to perform the test. The standard deviations in the auditory inspection time test were lower than those reported by previous workers (Deary, 1980; Irwin, 1984; Nettelbeck, Edwards and Vreugdenhil, 1986). This is probably, in part, due to pretesting for ability to perform the pitch discrimination required in the auditory inspection time test. Subjects were required to attain at least 90% performance on 19 items at long stimulus durations. Those subjects unable to meet this criterion were distinguished, in both the undergraduate sample and among the schoolchildren, by their low pitch discrimination scores on standard pitch discrimination tests. The fact that over one third of the undergraduates and almost one half of the schoolchildren had to be excluded from the analysis is inconvenient in terms of subject numbers in the data analyses, but might be considered to be analogous to ensuring that all subjects in, say, the visual IT test have adequate visual acuity. It is likely that some of the subjects who were labelled as being unable to perform the auditory inspection time discrimination had inspection times at or beyond 200ms. The consequence of this possibility would be a

lowering of the correlations between AIT and cognitive ability scores because of an attenuation of AIT range. It was considered to be more important to exclude all of those subjects who were not clearly capable of making the discrimination involved in the AIT test.

If a subject had been experiencing difficulty making the pitch discrimination involved in the auditory inspection time task, but had been, nevertheless, allowed to complete it, then by analogy with the visual IT test, this would have lead to the introduction of variance from the 'noise' parameter as well as the 'inspection time' parameter in perception (Vickers, Nettelbeck and Willson, 1972). In IT testing the stimuli must be easily discriminable in order that the difference between the discriminanda is well beyond the 'noise' level for subjects included in the analyses. Pitch discrimination skill was not a perfect discriminator of those subjects able or unable to complete the AIT task; there was a slight overlap in the pitch discrimination distributions of those who could and those who could not perform the auditory inspection time task.

In undergraduates pitch discrimination ability was not correlated significantly with intelligence, and pitch discrimination was the only variable which distinguished those subjects who could not perform the auditory inspection time test. Also, the correlation between auditory inspection time and pitch discrimination was non-significant. The schoolchildren had very similar mean AITs when compared with the undergraduates, but a smaller percentage of them were able to perform the auditory inspection time test. It is possible that, in undergraduates, pitch discrimination acts in a threshold manner such that above a critical level of pitch discrimination ability there is no or very little additional advantage in performing auditory inspection time. Also, it would be in keeping with the results to hypothesise that, whatever small overlap exists between auditory inspection time and pitch discrimination does not coincide with the variance shared between auditory inspection time and intelligence. In children it appeared that intelligence was a slightly



stronger correlate of pitch discrimination, in agreement with the study conducted by Lynn, Wilson and Gault (1989). However, a definitive statement about the differences in correlations between the two groups is not possible because, although the IQ-pitch discrimination correlation was significant in one case in the schoolchildren sample, the difference in correlation size between the schoolchildren and the undergraduates was not sufficiently large to be significant. Additionally, in the sample of schoolchildren, those subjects who could not perform the auditory inspection time were distinguished by their lower IQs as well as their poorer pitch discrimination ability.

This study offered evidence from two separate samples with different ages about 'what it means to be high-verbal' (Hunt, Lunneborg and Lewis, 1975). It indicated that the auditory inspection time-IQ correlation is higher in tests of verbal IQ rather than in non-verbal or mathematical tests. Also, the high-verbal schoolchildren had a pitch discrimination advantage, but no such pitch discrimination-verbal ability correlation was found in the undergraduate sample. It appeared, then, that advantages in two simple auditory abilities - those of auditory processing speed and of pitch discrimination - correlated significantly and somewhat selectively with verbal ability scores in schoolchildren, in agreement with the findings of Tallal (1989). It should be recalled, however, that the results of Chapter 4 were not so straightforward. In that study, the correlations of the Mill Hill Vocabulary test and Advanced Progressive Matrices with auditory inspection time might have led to a similar conclusion. However, the correlations between auditory inspection time and the verbal-numerical and non-verbal portions of the Alice Heim 5 test ran in the opposite direction, with the non-verbal test having a higher correlation with auditory inspection time. The study with AIT in children, therefore, is congruent with Nettelbeck's hypothesis (1987; Nettelbeck and Young, 1990) that IT has a stronger association with Verbal IQ than with Performance IQ in children. On the other hand, the suggestion that AIT might be more strongly related to Verbal IQ in adults would place it at odds with the results of visual IT studies, where there is generally a stronger

association with Performance IQ (Nettelbeck, 1987; Kranzler and Jensen, 1989; Deary, in press).

## **Chapter Six**

### **Auditory inspection time, cognitive ability and the 'Raz' task - I**

#### **6.1 Introduction**

##### **6.1.1 Auditory inspection time, pitch discrimination and intelligence**

Chapter 5 found little evidence to indicate that the AIT-cognitive ability test score association was accounted for by individual differences in pitch discrimination as indexed by the Seashore pitch test. However, three studies by Raz, Willerman and Yama (1987) raised again the issue of the part played by individual differences pitch discrimination ability in determining the correlation between auditory information processing tasks and cognitive ability test scores.

Using an adaptive staircase procedure, Raz, Willerman and Yama (1987) presented subjects with an auditory task consisting of two 20ms tones which were 850ms apart and unmasked. Subjects were asked to indicate whether the high tone came first or second. An adaptive algorithm (a modification of Levitt's (1971) threshold computation procedure) sought the smallest pitch difference where subjects could make accurate discriminations. Each run of the task was completed when 14 reversals were accumulated. The mean of the six best levels, i.e. those which represented the smallest pitch differences between the target tones, was used as the threshold. At the outset, the two tones were 100Hz apart and, for some subjects, this had to be reduced to 2 to 3 Hz in order to estimate their thresholds. Two conditions were introduced to the task. Two different ramp values (1ms and 9ms) were introduced to assess the effect of signal energy spectrum - steeper ramps have wider energy spectra - on subjects' performances on this task. For the sake of clarity, from now on, this pitch discrimination threshold task will be called the 'Raz' task.

The thresholds obtained from subjects completing the task had a skewed distribution, even after logarithmic transformation. IQ was correlated with frequency discrimination ability at -0.47 and -0.54 for the logarithmic transformations of the 1ms and 9ms-ramp tasks, respectively ( $n=25$ , both  $p<0.05$ ). The correlation was not caused by individual differences in practice, stimulus spectral composition (i.e. ramp duration), musical experience or demographic characteristics. A second experiment in the same report (Raz, Willerman and Yama, 1987) replicated the above mentioned relationship; using a signal ramp of 5ms, the correlation between logarithmic transformations of thresholds obtained by subjects on the Raz task and Cattell IQ scores was -0.52 ( $p<0.05$ ). However, when a more conventional method of estimating the auditory threshold was applied, the correlation fell to -0.36 ( $p<0.1$ ). Partialling out the effects of musical training had a negligible effect on the correlations.

In their third experiment, Raz Willerman and Yama (1987) devised a signal detection task in order to test the hypothesis that higher IQ subjects perform better on any novel, 'non-entrenched' task (Sternberg, 1981). After a warning, subjects attended to two 550ms 'observation periods' which were separated by a 500ms pause. In one of the observation periods, a 1000 Hz tone burst with a signal duration of 20ms and 5ms linear ramps was played. Again, an adaptive procedure was used to estimate detection thresholds, and the increment/decrement value was 1 dB. Signal detection thresholds were estimated in three conditions: in a condition using a continuous broad band noise masker, in a condition using a masker noise which had a spectral notch centred on the signal frequency of 1000 Hz, and in a non-masked quiet condition. IQ correlated at 0.16, 0.17 and -0.06 with signal detection thresholds ( $n=25$ ; all correlations non-significant and the first two in the opposite to the expected direction). They concluded that, whereas signal *recognition* correlated with ability, signal *detection* did not, though they recognised that their sample was too small to accept a null hypothesis with confidence.

Raz, Willerman and Yama (1987) concluded that the high IQ subject does not perform better on *any* novel 'non-entrenched' task. Moreover, they concluded that their results offered contrary evidence to those who reckoned that it was the high IQ subject's fast adaptation<sup>in</sup> the laboratory situation, increased motivation or lower distractibility that led to the association between simple processing tasks and IQ scores. As was discussed in Chapter 2, a related hypothesis was put forward by Irwin (1984) to explain the AIT-cognitive ability association in terms of the challenge and anxiety produced by novel tasks. However, note that this hypothesis is somewhat in opposition of that put forward by Mackintosh (1986), who suggested that the high IQ subject performs better given the *boring* nature of laboratory tasks, such as IT. While it might be true that bright subjects are more relaxed when faced with challenging tasks (Irwin, 1984) *and* more stimulated by boring tasks, there is clearly no consensus about the category to which AIT-type tasks belong. Attempts to explain the IT-IQ correlation in cognitive terms appear to be fairly flexible.

#### **6.1.2 Separating processing speed and discrimination ability**

The IT studies described in Chapters 4 and 5 provided evidence to indicate that indices derived from tasks that were devised to measure visual and auditory processing speed correlate at moderate levels with psychometric measures of intelligence. Also in support of the speed of processing hypothesis is the finding that visual IT and AIT have a moderate intercorrelation (Nettelbeck, Edwards and Vreugdenhil, 1986; Chapter 4 in this thesis; though this was not found by Irwin, 1984), i.e. the tasks appear to share a common requirement for speeded information processing but not a common element of discrimination. Evoked potential studies of AIT have suggested that there was perceptual speed-associated variance and task specific variance in IT tasks, and that the two sources of variance have separate correlates. The general speed factor appeared to be related to IQ, whereas the more task-specific variance was correlated with evoked potential indices of the early stages of information intake or pattern recognition (P200 rise time particularly) (Zhang, Caryl and Deary, 1989a,b; Deary and Caryl, 1988; Hall, 1988).

Although the studies described in Chapters 4 and 5 attempted to reduce the involvement of pitch discrimination as a potentially confounding variable in the AIT task, the results of Raz, Willerman and Yama (1987) indicated that, with briefly presented auditory stimuli, high IQ subjects made finer pitch discrimination judgements. Therefore, one must be wary of concluding prematurely that speed of processing is the key variable in auditory tasks, even in tasks which appear to index speed of processing fairly unequivocally. Raz, Willerman and Yama (1987) concluded from their results that,

...the frequency discrimination task imposes few time constraint[s] on information processing, calling only for fine perceptual resolution, with practically unlimited response time.

Raz, Willerman and Yama (1987) argued that the correlations between inspection time and cognitive ability might be explained by the higher IQ subject having: faster feature extraction; better sensory representation of stimuli; faster decision time; less bias in responding; or a combination of the above. They concluded that the decision time and response bias differences were ruled out by Nettelbeck and Lally (1976) and Lally and Nettelbeck (1977). Therefore, the remaining competitor to the suggestion that it is speed of information processing that underlies the IT-cognitive ability correlation is the hypothesis which states that the high IQ subject might have a better representation of sensory stimuli. Raz, Willerman and Yama (1987) commented that the separation of the 'processing speed', which they had opted for as the explanatory variable in auditory processing task-cognitive ability association studies in previous reports (Raz, Willerman, Ingmundson and Hanlon, 1983; Raz and Willerman, 1985), and 'fidelity of stimulus representation' explanations for the IT-cognitive ability association was not possible within the framework of the backward masking paradigm. Nevertheless, whereas they preferred the 'fidelity of stimulus representation' hypothesis,

Nettelbeck and his associates sided with the speed-of-processing explanation, assuming apparently that stimuli in their studies were too simple to produce a wide range of individual differences in feature representation. The possibility of



individual differences in fidelity of stimulus representation, however, has never been tackled directly.

The above-mentioned allegiance of Raz, Willerman and Yama (1987) to the 'fidelity of stimulus representation' hypothesis of the IT-cognitive ability association did not imply that they were denying the existence of faster mental speed *at some level* in the brighter individual. However, their opinion was that mental speed was a *higher order* manifestation of the *lower level* ability to make faithful stimulus representations,

The resolution of a system and its rate of information processing are intimately related. Under time constraints the system with lower need for external signal redundancy will respond faster than a noisy system, but this does not imply that the system actually processes information at a faster rate in terms of signal transmission velocity. Given the results of our experiments, the quality of the signal representation rather than speed of processing may be the key feature of an intelligent brain.

Recall that the accumulator model of visual perception articulated by Vickers, Nettelbeck and Willson (1972), stated that individuals, when making decisions about a stimulus in a two-choice decision task, must accumulate evidence for the two options against a background of noise. Evidence accumulates by the subject making inspections of the stimulus, each inspection takes a minimum amount of time, and a decision is made when the evidence for one of the alternatives passes a threshold. In the IT task, the stimulus presentation time is manipulated and, at very brief durations, it is assumed that the subject has not been able to accumulate sufficient evidence from the stimulus to make correct decisions reliably. However, for any given stimulus duration, because of individual differences in IT, some subjects will make more inspections of the stimulus than others, and a subject with a very short IT may achieve a faithful representation of a brief stimulus, whereas a subject with a long IT will have a poor representation.

Therefore, a fast IT might cause better representation of stimuli, and the finding of Raz, Willerman and Yama (1987) that high IQ subjects make better pitch discriminations to

briefly presented tone pairs may be explained by the fact that, given the constant, brief presentation time used in their studies I and II, as the pitch discrimination becomes more difficult, more inspections of the stimuli need to be taken in order to make reliable discriminations. This is unlikely to be a factor in pitch discrimination tasks such as the Seashore test, where the stimuli are presented for 500ms (which is far longer than any AITs reported in the above studies) and are unmasked. However, with 20ms tones of very similar pitch, even when they are unmasked and separated by 850ms, it is hypothesised that the subject with superior IT, i.e. the subject requiring shorter stimulus duration at any given frequency difference in order to reach criterion-level discrimination accuracy, will have an advantage.

In summary, the argument which pits processing speed against fidelity of stimulus representation might be a non-argument. The experiments reported so far in this thesis and those reported by Raz, Willerman and Yama (1987) are congruent with the fidelity of stimulus representation explanation *and* with an explanation which states that a fast inspection time is primarily an advantage in some basic information processing speed, i.e. the accumulator model of Vickers (Vickers, Nettelbeck and Willson, 1972; Vickers and Smith, 1986) may be extended to state that the person with a short IT achieves a more faithful representation of briefly presented stimuli.

### **6.1.3 Aims of the present study**

It was argued above that the Raz task might index speed of information processing abilities principally, or in addition to pitch discrimination ability. However, this hypothesis and the hypotheses of Raz, Willerman and Yama (1987) have been formulated in the absence of any study which has examined the associations between Raz and AIT empirically. As a result of the above considerations the aims of the studies in the present and the next Chapters were as follows,

- 1) An assessment was made of the correlation among the AIT, Raz and Seashore pitch

tasks. It was hypothesised that, because the Seashore task was unambiguously assessing pitch discrimination in an unspeeded fashion and because AIT was designed to assess speed of auditory processing, the Raz task would have a stronger association with Seashore pitch than would the AIT task (Chapter 6). It was also hypothesised that factor analysis of scores on these three auditory tasks would separate the AIT and Seashore tasks clearly, with Raz having an intermediate status (Chapter 7).

2) Because the brief stimulus durations used in the Raz task made it possible that the Raz task was assessing information processing speed to a significant degree, it was decided to construct two versions of the Raz task, with the stimulus tones played for 20ms and 160ms, respectively. It was hypothesised that the version of the Raz task with the briefer tone duration (20ms) would lead to higher estimates of pitch discrimination thresholds than would the version of the task with longer tones (160ms) (Chapter 6). It was also hypothesised that the version of the Raz task with briefer tone duration (20ms) would have a stronger association with AIT than with Seashore pitch and that the reverse would hold for the version with the longer tone duration (160ms) (Chapter 6).

3) Attempts were made to replicate the correlations between AIT and Raz tasks and cognitive ability test scores, and to observe any changes in these correlations when individual differences in Seashore pitch discrimination scores were controlled for (Chapters 6 and 7).

4) An attempt was made to test the hypothesis that the correlation between cognitive ability test scores and simple tests of auditory processing are higher at subjects' first meeting with the latter tasks, as suggested by Irwin (1984) and Sternberg (1981) (Chapter 6).

## **6.2 Method**

### **6.2.1 Subjects**

Twenty-eight third year psychology undergraduates were recruited for the study. There were 18 women and 10 men, aged 20 to 23. No subjects reported having hearing deficits.

### **6.2.2 Cognitive ability tests**

#### **Alice Heim 6 test (AH6)**

This is a cognitive ability test designed to discriminate among individuals with high cognitive ability levels, such as university undergraduates. There are 18 example items and 60 test items. Only the test items were used to derive scores. Subjects completed the example items with feedback on the correctness of their performance and, if they were unsure how the correct answer was arrived at, it was explained to them. The 'AG' version of the AH6 was used, which resulted in three scores. The Verbal score was the number of correct verbal reasoning- and vocabulary-type items that were solved correctly. The Numerical and Diagrammatic score was derived from items that involved reasoning about arithmetical problems and geometric forms. The total score was the sum of the Verbal and Numerical and Diagrammatic scores. The AH6 was administered as a group test. There was no time limit for the examples, and 40 minutes was allowed to complete the 60 test items which consisted of a mixture of the three types of item.

#### **Raven's Advanced Progressive Matrices**

This test was described in Chapter 4. Set I was used for practice, and the number of correct Set II items was used as the total score. Administration was as described in Chapter 4.

### **6.2.3 Auditory tests**

#### **Auditory Inspection Time (AIT)**

In order to make the AIT test more similar to the RAZ test, which is the main focus of this

Chapter, a slightly revised form of AIT test was constructed for this study. The stimuli to be discriminated were square wave tones of 770 Hz and 880 Hz. Stimulus tones were played at 80 dB with instantaneous rise and fall times. These were cued by a sound which was a mixture of the two stimulus tones, and which lasted for 300ms. The time between cue offset and stimulus tone onset was 700ms. The two stimulus tones, the order of which was to be discriminated (i.e. 'High-Low' or 'Low-High') were played consecutively, with no gap between them. The stimulus tones were backward masked with a chord which was a mixture of the two stimulus tones and an intermediate tone of 825 Hz, all of which were played simultaneously. The backward mask lasted for 700ms. There was no stimulus-mask gap.

The psychophysical procedure used to derive the subjects' auditory inspection times was the parameter estimation by sequential testing (PEST) technique of Taylor and Creelman (1967). This was described for the visual IT tests in Chapter 4. The procedure used in this study for the estimation of auditory inspection times was similar. The starting stimulus duration was 200ms (for *each* of the two stimulus tones). The stimulus duration change was halved with each reversal of the staircase. The test was stopped when the algorithm attempted to change from 2 to 1 ms. The rules for changing duration were as described in Chapter 4 and were applied to the auditory test as follows. The initial change of duration in the staircase was 75ms. If more than one error was committed in a block of five trials, the tone duration was increased for the next block. If one error was made in the block of five trials, more trials were given. If an error was committed in any of the next five trials, the duration was increased but, if no further errors were made at that duration, i.e. if the subject then had nine out of 10 correct, the duration was reduced. The subject's inspection time represented the algorithm's estimate of the duration at which the subject was 85% correct in his or her responses.

The stimulus presentation and the algorithm were controlled by a BBC B microcomputer. Stimuli were played on the microcomputer's internal loudspeaker. Subjects were

encouraged to respond at their leisure and to attempt to maintain maximum accuracy. Responses were made on the computer's keyboard, with the 'Z' key representing the stimulus tone sequence 'Low-High' and the '/' key representing the sequence 'High-Low'. These keys are placed conveniently for use by the index fingers of the left and right hands, respectively. Subjects wrote out a card with the response modes and placed these in front of them while undertaking the task. A response to an item initiated the next item with an intertrial interval of between one and two seconds. The computer screen offered instructions which supplemented detailed verbal instructions given by the author.

### **RAZ test**

The RAZ test was designed to be as similar as possible to that used by Raz, Willerman and Yama (1987). This test involved subjects in discriminating the temporal order of two tones which differed in pitch. The tones were 80 dB square wave tones with instantaneous rise and fall times. By contrast with the auditory inspection time task, the tones remained fixed in duration, but were varied with respect to their pitch difference. The tones for the test were generated by a BBC B microcomputer and played using the amplifier and loudspeaker of a UHER reel-to-reel tape recorder which was interfaced with the microcomputer. Two forms of the Raz test were devised. In the 'RAZshort' test the duration of the two tones, the order of which was to be discriminated (i.e. 'High-Low' or 'Low-High'), was 20ms for each tone. In the 'RAZlong' test the stimulus duration of each tone was fixed at 160ms.

A visual cue - 'READY' then 'LISTEN' printed consecutively on the computer monitor in front of each subject - preceded each tone pair by 1000ms. Tones lasted for either 20ms or 160ms and were separated by a silent gap of 860ms. There was no backward mask. Subjects responded to the tone pairs, indicating whether they thought the order was 'High-Low' or 'Low-High' by pressing the right or left button of a two-choice response box which was interfaced with the microcomputer which controlled the task. As with all versions of the AIT task, the left button was used for the 'Low-High' response and the



right button for the 'High-Low' response. Subjects were encouraged to respond at their leisure and to attempt to maintain maximum accuracy in their responses.

A PEST adaptive staircase psychophysical procedure was used to alter the *pitch difference* between the tones. The tones began the test at 770 Hz and 880 Hz for the high and low tones, respectively. Thereafter the PEST procedure was, *mutatis mutandis*, similar to that used for the auditory inspection time task. The pitch difference was increased or decreased according to the subject's prior performance record. The minimum pitch discrimination ability which the system was able to detect in subjects was 3 Hz, owing to limitations in the tone-generating procedure of the BBC B microcomputer. Subjects finished the task when the PEST procedure had found the point at which the subjects were 85% correct in discriminating the pitch of the two tones. Therefore, for all subjects, the final tones that they heard were  $825+(d/2)$  Hz and  $825-(d/2)$  Hz, where  $d$  represented the pitch difference at which they were 85% accurate.

Stimulus tones and the PEST procedure were controlled by a BBC B microcomputer.

### **Seashore pitch test**

This 50-item test of pitch discrimination was described in Chapter 5 and was administered in the same way to all subjects in the present study.

#### **6.2.4 Procedure**

All subjects attended for 3-hour sessions on Fridays in three consecutive weeks, i.e. the total testing time for each subject was nine hours. Half of the subjects attended for morning sessions and the others attended in the afternoons.

In the first session subjects undertook the Alice Heim 6 test as described above. This was done as a group test in a quiet basement room under examination conditions. After a break of about 15 minutes they undertook the Seashore pitch test. This took place in individual

basement cubicles. Subjects heard the instructions and test items through headphones and made the appropriate responses on the standard Seashore response sheet. Instructions were then given for the auditory inspection time task. This was done as a group and subjects then went into their individual cubicles to complete the task. Unlike the previous AIT task, which was run in a group setting, individuals undertook this PEST-controlled form of the AIT task at their own pace.

In the second session subjects began by completing the Raven's Advanced Progressive Matrices tests under examination conditions as described above and in Chapter 4. After a 15-minute break, subjects were given detailed instructions for the completion of the RAZ tasks. They then went into individual basement cubicles and completed the RAZshort (20ms tone duration) task. A further break of 10 minutes was followed by the completion of the RAZlong (160ms tone duration) task.

In the third session subjects undertook the AIT and Seashore tasks for a second time, using the same settings and procedures as in session one. After a break of 15 minutes they undertook the RAZlong test for a second time, which was followed by the RAZshort test after a further break of 10 minutes. Thus, the occurrence of the RAZshort and RAZlong tests was counterbalanced across the weeks. Because the PEST algorithm leads to an estimate that is tailored to each subject's performance, it was inevitable that some subjects finished the AIT and RAZ tasks sooner than others. Therefore, the times given for breaks represent the minimum time that people had for a rest between tests.

## **6.3 Results**

### **6.3.1 Descriptive statistics and group comparisons**

Table 6.1 shows the means and standard deviations for all the tests undertaken by the subjects in the study. Data were complete for all tests except the AIT task. Seven of the 28 subjects did not provide AIT data. These subjects were unable to perform the

discrimination reliably enough at any duration (the maximum allowed duration of the staircase was 248ms) to enable them to complete the PEST procedure with the AIT stimuli. The median AIT score of subjects improved significantly over the fortnight that separated the two test occasions ( $Z=2.67$ ,  $p<0.01$ ). The Seashore test also showed a significant practice effect ( $p<0.01$ ; Table 6.1)

The RAZ test containing the 20ms tones (RAZshort) improved significantly over the week that separated the two task occasions ( $Z=3.0$ ,  $p<0.01$ ), whereas there was no significant improvement in the RAZ task with the 160ms tones (RAZlong;  $Z=0.43$ , ns). Little weight should be placed upon these results because they are confounded by the effects of practice across the two forms of the RAZ task; the RAZshort test was the first RAZ test in session two and the second RAZ test in session three. The mean of the RAZshort and RAZlong tasks across the two weeks for all subjects allows a comparison of the effects of tone duration on the RAZ task which is counterbalanced for the effects of practice. The mean of the two RAZshort tasks was 11.8 Hz (SD 15.0, range 3 to 59.5). The mean of the two RAZlong tasks was 6.7 Hz (SD 7.9, range 3 to 40.5). The difference between the two forms of the RAZ task was highly significant (Wilcoxon signed ranks test  $Z$  score=4.05,  $p<0.01$ ). Therefore, the duration of the stimulus tones had a significant effect on the pitch duration thresholds obtained from the RAZ task.

### **6.3.2 Correlations among cognitive and auditory tests**

Table 6.2 shows the cognitive test score intercorrelations. All correlations were significant at  $p<0.01$ , with the Raven APM-AH6 correlations at 0.55 and 0.74 for the Verbal and Numerical-Diagrammatic subscales, respectively. The two subscores of the AH6 test correlated at 0.74.

Table 6.3 shows the auditory test score intercorrelations. Table 6.4 shows the same test intercorrelations performed on the logarithmic transformations of the AIT and RAZ test

**Table 6.1**

Descriptive statistics (mean and SD) and first versus second test comparisons for tests used in the present study (n=28, except for AIT tests where n=21).

	First test	Second test	Statistical test result	p
AIT <sup>a</sup> (ms)	87.5 (57.8)	76.6 (46.6)	2.93 <sup>e</sup>	<0.01
RAZ-short (Hz)	14.6 (19.9)	8.9 (12.9)	2.06 <sup>e</sup>	<0.01
RAZ-long (Hz)	8.1 (13.1)	5.3 (3.7)	0.43 <sup>e</sup>	ns
Seashore (no. correct)	40.5 (6.0)	42.9 (8.0)	2.95 <sup>f</sup>	<0.01
Raven APM <sup>b</sup>	27.9 (4.5)	-		
AH6 <sup>c</sup> Verbal	17.4 (5.7)	-		
AH6 N&D <sup>d</sup>	13.5 (5.2)	-		
AH6 Total	30.9 (10.2)	-		

<sup>a</sup>Auditory inspection time

<sup>b</sup>Advanced Progressive Matrices

<sup>c</sup>Alice Heim 6 test

<sup>d</sup>Numerical and diagrammatic

<sup>e</sup>Z score from Wilcoxon Signed Ranks test

<sup>f</sup>Student's paired t test

**Table 6.2**

Correlations among cognitive ability tests (n=28).

	Raven APM	AH6 Verbal	AH6 N&D	AH6 Total
Raven Advanced Progressive Matrices	-	0.55**	0.74***	0.69***
AH6 <sup>a</sup> Verbal		-	0.74***	0.94 <sup>c</sup> ***
AH6 N&D <sup>b</sup>			-	0.92 <sup>c</sup> ***

\*\* p<0.01; \*\*\* p<0.001

<sup>a</sup>Alice Heim 6 test

<sup>b</sup>Numerical and diagrammatic

<sup>c</sup>Subtests correlated with total that contains the subtest

**Table 6.3**

Correlations among auditory tests (n=28, except for correlations involving AIT, where n=21). See key at bottom of page for test name abbreviations.

	AIT1	AIT2	RAZsh1	RAZsh2	RAZlo1	RAZlo2	Sea1	Sea2
AIT1	-	0.96***	0.43*	0.37+	0.39+	0.46*	-0.45*	-0.31
AIT2		-	0.40+	0.39+	0.31	0.45*	-0.47*	-0.24
RAZsh1			-	0.67***	0.76***	0.53**	-0.63***	-0.51**
RAZsh2				-	0.91***	0.86***	-0.66***	-0.42*
RAZlo1					-	0.64***	-0.54**	-0.29
RAZlo2						-	-0.74***	-0.61***
Sea1							-	0.84***

**Table 6.4**

Correlations among auditory tests after logarithmic transformation of the Auditory Inspection Time and RAZ test scores (n=28, except for correlations involving AIT where n=21). See key at bottom of page for test name abbreviations.

	log AIT1	log AIT2	log RAZsh1	log RAZsh2	log RAZlo1	log RAZlo2	Sea1	Sea2
log AIT1	-	0.94***	0.50*	0.30	0.33	0.42+	-0.45*	-0.30
log AIT2		-	0.58**	0.42+	0.33	0.46*	-0.53*	-0.30
log RAZsh1			-	0.81***	0.78***	0.66***	-0.77***	-0.63***
log RAZsh2				-	0.80***	0.76***	-0.77**	-0.57**
log RAZlo1					-	0.59**	-0.61***	-0.43*
log RAZlo2						-	-0.76***	-0.66***

+ p<0.1; \* p<0.05; \*\* p<0.01; \*\*\* p<0.001

AIT1 and AIT2: first and second tests of Auditory Inspection Time, respectively.  
 RAZsh1 and RAZsh2: first and second tests, respectively, of the RAZ test involving a stimulus duration of 20ms.  
 RAZlo1 and RAZlo2: first and second tests, respectively, of the RAZ test involving a stimulus duration of 160ms  
 Sea1 and Sea2: first and second tests, respectively of the Seashore pitch test.

scores, but not for the correlations between the two administrations of the Seashore pitch test. These transformations were performed because the raw data were not normally distributed. Transformed AIT scores showed acceptable distributions, whereas the RAZ scores were still somewhat skewed. The reliability of the AIT test over a period of two weeks was very high, over 0.9, and the Seashore test-retest reliability was over 0.8. The RAZlong and RAZshort test reliabilities were 0.64 (0.59 for the transformed score) and 0.67 (0.81 for the transformed score), respectively. The average correlation between transformed RAZlong and RAZshort test scores was 0.75.

AIT from the first and second test, respectively, correlated at -0.45 and -0.47 with the first Seashore test. The correlations fell to non-significant levels with the second Seashore test. The RAZ test correlations were also higher with the first Seashore test, generally above -0.7 for the transformed RAZ scores. The average correlation between transformed RAZ scores and first Seashore test scores was 0.73, and between transformed AIT scores and the first Seashore test was 0.49, indicating that, compared with AIT, RAZ had a stronger association with pitch discrimination.

It was hypothesised that the original RAZ test which used 20ms tones was, in part, a speed of auditory processing test, whereas it was thought that the version of the test which had 160ms tones would be more strongly associated with pitch discrimination than with speed of processing. This led to the hypotheses that the RAZshort would be more strongly correlated with the AIT test than would the RAZlong test, and that the magnitudes of the RAZ correlations with the Seashore pitch test would be in the opposite direction. The mean correlation between transformed RAZshort and transformed AIT scores was 0.45 versus 0.38 for the RAZlong test scores. This was in the expected direction, but represented a small difference that was not taken to be significantly different. The mean correlation between transformed RAZshort and Seashore test scores was 0.68 versus 0.61 for the RAZlong test scores. This was in the direction opposite to that expected from the hypotheses, but the small difference was not significant.



Table 6.5 shows the partial correlations between AIT and RAZ test scores when the scores from the first Seashore test were controlled for. The correlations remained positive, but almost all fell to non-significant levels, indicating that a substantial amount of the variance that is shared by the RAZ and the AIT tests was accounted for by scores on the the Seashore pitch test. The highest partial correlations were between the logarithmic transformations of the Raz thresholds from the first RAZshort test and AIT threshold estimates.

#### **6.3.4 Ability and auditory test intercorrelations**

Table 6.6 shows the correlations between cognitive ability test scores and the auditory test scores. Table 6.7 has the correlations between cognitive test scores and the transformed AIT and RAZ test scores. Correlations between AIT scores and ability test scores were all in the expected direction, but were non-significant, with only the AIT-AH6 numerical and diagrammatic and total scores showing a trend beyond the 0.1 level of significance. The correlations between AIT and Raven scores were between 0.1 and 0.2. The correlations between RAZ test scores and AH6 scores were generally very low and all were non-significant. The RAZ-AH6 verbal correlations were low and positive (i.e. in the opposite direction to that expected from past research), whereas the RAZ-AH6 numerical and diagrammatic correlations were small and negative. The RAZ-Raven correlations were slightly higher and, before the RAZ scores were transformed, two of the four correlations met the 0.1 level of significance. The mean RAZshort-Raven correlation was 0.18 and for the RAZlong test was 0.30. Seashore test scores showed small, non-significant correlations with the cognitive test scores. The correlations with both of the non-verbal tests used (Raven and AH6 numerical and diagrammatic) were positive, with a mean of 0.20, and with the AH6 verbal test they were near to zero and negative.

#### **6.3.5 Partial correlations between AIT and ability tests**

Table 6.8 shows the partial correlations between AIT scores and RAZ test threshold estimates and cognitive ability test scores when the effects of Seashore pitch were

**Table 6.5**

Partial correlations between Auditory Inspection Time (AIT) and Raz test scores, and their logarithmic transformations, controlling for Seashore pitch test scores (n=21). For key to variable names see bottom of page.

	AIT1	AIT2	logAIT1	logAIT2
RAZshort1	0.22	0.16	0.18	0.13
RAZshort2	0.11	0.10	0.05	0.07
RAZshortMEAN	0.20	0.15	0.16	0.12
log RAZshort1	0.35	0.37+	0.28	0.35
log RAZ short2	0.07	0.14	0.01	0.11
RAZlong1	0.19	0.06	0.14	0.05
RAZlong2	0.21	0.18	0.16	0.12
RAZlongMEAN	0.22	0.15	0.17	0.10
log RAZlong1	0.18	0.08	0.13	0.07
log RAZlong2	0.16	0.15	0.13	0.10

+ p<0.1

AIT1 and AIT2: first and second tests of Auditory Inspection Time, respectively.

RAZshort1 and RAZshort2: first and second tests, respectively, of the RAZ test involving a stimulus duration of 20ms.

RAZlong1 and RAZlong2: first and second tests, respectively, of the RAZ test involving a stimulus duration of 160ms.

**Table 6.6**

Correlations between cognitive ability tests and auditory tests (n=28, except for correlations involving AIT where n=21). See bottom of page for key to auditory test variable names.

	Raven Advanced Progressive Matrices	Alice Heim 6 Verbal	Alice Heim 6 Numerical and Diagramatic	Alice Heim 6 Total
AIT1	-0.14	-0.23	-0.34	-0.29
AIT2	-0.14	-0.30	-0.38+	-0.36+
RAZshort1	-0.12	0.17	-0.05	0.07
RAZshort2	-0.33+	0.10	-0.25	-0.07
RAZlong1	-0.34+	0.15	-0.15	0.01
RAZlong2	-0.27	0.05	-0.20	-0.07
Seashore1	0.16	-0.09	0.28	0.09
Seashore2	0.18	-0.06	0.19	0.06

+  $p < 0.1$

AIT1 and AIT2: first and second tests of Auditory Inspection Time, respectively.

RAZshort1 and RAZshort2: first and second tests, respectively, of the RAZ test involving a stimulus duration of 20ms.

RAZlong1 and RAZlong2: first and second tests, respectively, of the RAZ test involving a stimulus duration of 160ms

Seashore1 and Seashore2: first and second tests, respectively of the Seashore pitch test.

**Table 6.7**

Correlations between cognitive ability tests and logarithmic transformations of auditory tests scores (n=28, except for correlations involving AIT where n=21). See bottom of page for key to auditory test variable names.

	Raven Advanced Progressive Matrices	Alice Heim 6 Verbal	Alice Heim 6 Numerical and Diagramatic	Alice Heim 6 Total
log AIT1	-0.10	-0.21	-0.31	-0.27
log AIT2	-0.13	-0.31	-0.40+	-0.37+
log RAZshort1	-0.14	0.11	-0.16	-0.02
log RAZshort2	-0.10	0.17	-0.11	0.04
log RAZlong1	-0.25	0.19	-0.07	0.07
log RAZlong2	-0.17	0.07	-0.12	-0.02

+ p<0.1

AIT1 and AIT2: first and second tests of Auditory Inspection Time, respectively.

RAZshort1 and RAZshort2: first and second tests, respectively, of the RAZ test involving a stimulus duration of 20ms.

RAZlong1 and RAZlong2: first and second tests, respectively, of the RAZ test involving a stimulus duration of 160ms.

**Table 6.8**

Partial correlations between cognitive ability tests and Auditory Inspection Time (AIT) tests and RAZ test scores and their logarithmic transformations controlling for Seashore pitch test scores (n=21 for correlations involving AIT; n=28 for correlations involving RAZ). The suffixes 1 and 2 refer to the first and second AIT and RAZ tests.

	Raven Advanced Progressive Matrices	Alice Heim 6 Verbal	Alice Heim 6 Numerical and Diagramatic	Alice Heim 6 Total
AIT1	-0.25	-0.29	-0.32	-0.32
AIT2	-0.25	-0.38+	-0.37+	-0.39+
log AIT1	-0.21	-0.27	-0.29	-0.29
log AIT2	-0.27	-0.41+	-0.39+	-0.42+
RAZshort1	-0.03	0.15	0.17	0.17
RAZshort2	-0.31+	0.06	-0.08	-0.01
RAZlong1	-0.31+	0.12	0.01	0.07
RAZlong2	-0.24	-0.02	0.08	-0.01
log RAZshort1	-0.03	0.06	0.08	0.08
log RAZshort2	0.03	0.17	0.17	0.18
log RAZlong1	-0.20	0.18	0.13	0.17
log RAZlong2	-0.07	0.00	0.15	0.08

+  $p < 0.1$

controlled for. The correlations between AH6 scores and the AIT first test scores, whether raw or transformed, remained in the region of 0.3. The correlations between AIT second test scores and AH6 scores were higher, in the region of 0.4, and all showed a statistical trend ( $p < 0.1$ ) in the expected direction. Correlations between AIT and Raven scores, when the effects of Seashore pitch were partialled out, ranged from -0.21 to -0.27 and all were non-significant. The partial correlations between raw Raz test scores and cognitive ability test scores were non-significant and fairly evenly distributed about zero except for two modestly sized correlations in the expected direction between Raven's Advanced Matrices and RAZlong1 and RAZshort2 (both -0.31,  $p < 0.1$ ). When the partial correlations were recomputed using the logarithmic transformations of the Raz scores there was even less evidence of a remaining significant association between Raz test scores and cognitive ability indices. The mean of the 12 partial correlations between the four transformed Raz test scores and the three independent cognitive ability scores was 0.06 (i.e. in the direction opposite to that expected), whereas the mean of the six partial correlations between the two estimates of AIT threshold and the three ability test scores was -0.31. Though this latter correlation does not reach a trend for the 21 subjects, it is of a similar magnitude to the size of correlation that might be expected in an UK undergraduate sample (see Chapters 4 and 5).

#### 6.4 Discussion

Seashore test scores and AIT threshold estimates proved to be reliable across two and one week periods, respectively, though both showed significant effects of practice across this time. Nettelbeck and Wilson (1985) also demonstrated an effect of practice across a two week period for visual IT. Raz tests showed moderately high reliability, and the two forms of the test showed intercorrelations as high as the individual test-retest reliabilities. Increasing the stimulus duration in the Raz test did have a significant effect on the pitch threshold achieved, which contradicts the view of Raz, Willerman and Yama (1987) that speed of processing is not important in the 20ms version of the test.



The effect of practice on the Raz tests could not be assessed separately for the two forms of the test because the type of Raz test was confounded with the order in which the tests were given. This aspect of the experiment's design was deliberate. The more important issue was the effect of stimulus duration on the Raz tests and their intercorrelation with cognitive ability test scores. Therefore, the stimulus duration and practice effects were counterbalanced across the two weeks. It would have been possible to have half of the subjects perform the Raz tasks in different orders, but that would have confounded individual differences with practice. Therefore it was felt that the present design served to demonstrate the effects of tone duration and individual differences satisfactorily.

As hypothesised in the introduction, Raz test thresholds did show higher correlations with Seashore test scores than did AIT threshold estimates, though both tests correlated significantly with Seashore. Therefore, even the AIT test, in this sample, showed some association with unspeeded pitch discrimination. It might be hypothesised that the higher Raz-Seashore correlations were the result of the fact that these correlations were based on the entire sample of 28, whereas the AIT-Seashore correlations were based on only 21 subjects, and that the 7 subjects missing from the AIT data led to an attenuated Seashore range and, therefore, smaller correlations. However, when the Raz-Seashore correlations were recomputed for the 21 subjects who were able to perform the AIT task, the mean Raz-Seashore correlation for the eight correlations between the two Seashore tests and the two forms of the Raz test each taken twice was -0.72 (range, -0.56 to -0.82). When the transformed Raz scores were used the results were similar, with a mean Raz-Seashore correlation of -0.69 (range, -0.53 to -0.83).

By the second time of testing on the Seashore test the AIT-Seashore correlations became non-significant in value. This may have been caused by the practice in the Seashore test leading to some ceiling effect and, as a result, lower correlations. Partial correlations between Raz and AIT thresholds controlling for Seashore pitch scores confirmed that the tasks probably share pitch discrimination-based variance. There was some remaining

positive correlation, even when pitch was controlled for, though the correlations were non-significant, due to the small sample size. The largest remaining correlation was between AIT and the first meeting of the RAZshort test, arguably the Raz test most loaded on speed of processing. Therefore, there is some evidence to indicate that AIT and Raz tests share variance that is pitch discrimination-related and some variance that does not relate to unspeeded pitch discrimination.

Correlations between the auditory tasks and cognitive ability tasks in the present study were unlikely to be highly significant owing to the relatively small number of subjects and their restricted ability range, though the number of subjects and their ability range was similar to that used by most other studies of auditory processing and intelligence (Nettelbeck, Edwards and Vreugdenhil, 1986; Raz, Willerman, Ingmundson and Hanlon, 1983; Raz and Willerman, 1985; Raz, Willerman and Yama, 1987). Correlations between AIT thresholds and cognitive ability test scores were all in the expected direction. As found in earlier chapters, the correlations with Raven's Advanced Matrices were low and non-significant. The mean correlation between AIT and AH6 Verbal score was -0.26 and with AH6 Numerical and Diagrammatic was -0.36. Therefore, in this study there was no evidence of superior association between AIT and verbal ability in adults. There were very small differences between the correlations of the first and second testings of AIT with cognitive ability test scores, and the correlations with the second test were higher. Correlations between AIT and cognitive ability scores changed minimally, but generally improved (Table 6.8), when the effects of Seashore pitch scores were controlled for, indicating that variance shared between AIT and cognitive ability was probably not related to pitch discrimination ability.

The mean correlation between the four Raz tests and Raven's Advanced matrices was 0.26, and two of these four correlations showed a trend in the expected direction; correlations with AH6 scores were smaller. Improvement of the distribution of scores had little effect on the AIT-cognitive ability correlations, but tended to reduce the Raz-cognitive

ability correlations to even lower levels, as did the procedure of partial correlation where Seashore scores were controlled. Though there were few correlations even tending toward significance, there was no evidence to indicate that Raz test thresholds tended to correlate better with cognitive test scores at the first meeting with the Raz test. Seashore pitch test scores correlated at about 0.2 with non-verbal ability test scores but, although these appeared to be congruent with the results reviewed in Chapter 3 and those of Lynn, Wilson and Gault (1989) and McLeish (1950), they fell short of conventional levels of significance.

It was remarked in the introduction that cognitive hypotheses formulated in an effort to 'explain away' (Sternberg 1988) the IT-cognitive ability association could be very flexible, with one hypothesis characterising IT tasks as challenging and anxiety-provoking (Irwin, 1984), whereas another hypothesis was based on the premise that the IT task presented a boring situation to subjects (Mackintosh, 1986). Therefore, in a spirit of balance, it should be remarked here that reductionistic explanations of the IT- and Raz task-cognitive ability association can be equally flexible. Raz, Willerman and Yama (1987) hypothesised that apparently faster information processing speed might be caused by the brighter subject's brain having greater 'hardware redundancy', i.e. more information might be gathered from briefly presented stimuli if redundant elements of the stimulus gathering apparatus in, say, the cochlea were connected in a parallel fashion. Therefore, according to Raz, Willerman and Yama (1987), biological differences in the ability to make accurate stimulus representations might cause apparent differences in speed of information processing. Contrariwise, it was possible to argue with equal ease that if the hypothesised accumulator apparatus of Vickers (Vickers, Nettelbeck and Willson, 1972) had a biological basis, then biological differences in speed of information processing might cause apparent differences in fidelity of stimulus representation.

In summary, the group comparison aspects of this study were successful, as were the attempts to discover the interrelationships among the auditory tasks. However, the

numbers proved too small to afford definite conclusions from the auditory test-cognitive ability associations. Therefore, to explore the latter further, a larger group will be tested in the next Chapter.

## **Chapter Seven**

### **Auditory inspection time, cognitive ability and the 'Raz' task - II**

#### **7.1 Introduction**

The principal aim of the present study was to investigate further the associations among AIT, Raz and Seashore test scores and their correlations with cognitive ability test scores. The study in Chapter 6 indicated that Seashore and AIT task scores shared a significant amount of variance, but suggested that this variance was not associated with cognitive ability. Chapter 6 indicated also that stimulus duration had a significant effect on Raz thresholds. The Raz and Seashore task scores also shared variance, but correlations between Raz and ability test scores were not high enough or consistent enough for the testing of hypotheses. There was some evidence to indicate that, as expected, the Raz test was more closely associated with unspeeded pitch discrimination than was the AIT test. Clearly, larger numbers of subjects were required to examine these issues further.

Until this point in the thesis, the relationships between the auditory tests and the cognitive ability tests used in the various studies have been examined by traditional, correlation-based methods. This is true, also, of all published work in the field of inspection time and intelligence (e.g. Kranzler and Jensen, 1989; Juhel, 1991). Raw correlations have been used to test for the amount of variance shared by two tests, and partial correlation has been used to control for the effects of confounding or mediating variables. In addition, various forms of factor analysis-type methods have been used to describe the latent variables that might underlie the various test intercorrelations (e.g. Nettelbeck and Young, 1990). This last method, especially, is an exploratory method and often puts the researcher in a position where results drive the hypotheses which are formed.

Arguably, a better state of affairs exists where hypotheses are set up explicitly in advance and are tested formally in the analyses for goodness-of-fit to the data. Therefore, in this Chapter, in addition to the conventional correlation, partial correlation and factor analytic methods that will be used to examine the interrelationships among the Raz, AIT, Seashore pitch and cognitive ability tests, a more hypotheses-testing approach will be introduced. Structural modelling of the test score relationships will be performed using a confirmatory factor analysis method (Bentler, 1989; Cuttance and Ecob, 1990). In this approach the hypothesised relationships among the test variables are expressed as a series of linear equations. The series of equations that express such relationships becomes the model under test, and various statistical procedures exist for assessing the goodness-of-fit of such models to the relationships among the test variables that is found in the test score covariance matrix.

## **7.2 Method**

### **7.2.1 Subjects**

The schoolboys and schoolgirls who were tested for Study 2 in Chapter 5 were recontacted two years later, by which time they had moved to separate secondary schools, for the study of AIT development to be reported in Chapter 8. Data were gathered during this follow-up session for the present study of the relationships among AIT, the RAZ test and Seashore pitch discrimination.

Of the 117 children who provided complete data sets for Chapter 5, 110 were traced and agreed to take part in the present study. The seven children who were not traced had moved with their parents from the Edinburgh area and the distances involved made retesting impracticable. Of the 110 children traced, two of them were absent from school, owing to illness, consistently over the period of the retesting and were, therefore, not



tested. Therefore, the sample of subjects for this study and the study to be reported in Chapter 8 was 108 second year secondary school children (92.3% of the original sample; 54 boys and 54 girls), with a mean age of 13.7 years (SD 0.38 years). All subjects had already taken part in a previous study of cognitive ability and AIT (Chapter 5, Study 2). Follow-up periods ranged from 23 to 25 months.

### **7.2.2 Cognitive ability tests**

#### **Raven's Standard Progressive Matrices (Raven, 1958)**

Raven's Standard Progressive Matrices was used as a test of non-verbal ability. The same test that had been used about two years earlier was readministered to the subjects. It was given as a 40 minute timed test after instruction on the principles behind the questions and instruction on the correct manner for completing the answer sheet.

#### **Mill Hill Vocabulary Test, Form 1 Senior Parts A and B (Raven, Raven and Court, 1982)**

Mill Hill Vocabulary Test, Form 1 Senior Parts A and B (i.e. synonyms and definitions) was used as a test of verbal ability. The choice of the particular test Form used here presented some difficulty. With the passing of two years it was inevitable that the schoolchildren's vocabularies would have improved, and it was felt that the use of the Junior Forms of the Mill Hill test might lead to a ceiling effect. Therefore, to accommodate the expected higher range of ability that was expected, and to preserve continuity with the testing of the subjects about two years previously, the Form 1 Senior version of the Mill Hill test was used. Both the multiple choice synonym subtest and the word definition subtest were used. The Mill Hill test was given as an untimed test under examination conditions.

### **7.2.3 Auditory tests**

#### **Auditory inspection time**

The test which had been administered to the same subjects about two years earlier was re-administered, i.e. this was the tape-recorded method of constant stimuli (descending series) version which was described in Chapters 4 and 5. Since the RAZ test was to be administered using a PEST-controlled adaptive staircase procedure, it might have been ideal to test AIT using the PEST-controlled AIT test described in Chapter 6. However, it was considered that the factor of primary importance was to test the children using the same AIT test that they had met previously in order to study AIT development in the next Chapter. The only difference between the AIT test as used in the present study and that given to the children two years earlier was the omission of the 6ms and 11ms stimulus durations from the taped AIT test. This was done to reduce the time of testing and because no adult or child subject had, until this time, achieved an auditory inspection time lower than 20ms. Therefore, it was felt that these durations were superfluous.

#### **Raz test (Raz, Willerman and Yama, 1987)**

The Raz test used in this study was the 20ms tone duration test used in the previous Chapter. Because there had been little difference between the 20ms and 160ms Raz tests in terms of their correlations with cognitive ability, AIT and Seashore pitch tests it was decided to use the version which was closest to that used by Raz, Willerman and Yama (1987). Therefore, as in the previous Chapter, this test involved subjects hearing two tones, each for 20ms and with a silent gap lasting 860 ms between them. The tones were unmasked and the cue was visual, as before. Tones were played at 80 dB and were generated by a BBC B microcomputer and played through the amplifier and loudspeakers of a UHER reel-to-reel tape recorder. A PEST adaptive staircase procedure, controlled by the same computer programme, altered the pitch difference between the tones according to subject's recent performance. The final score represented the pitch difference at which subjects were 85% correct in discriminating the difference between the tones, i.e.

responding 'High-Low' or 'Low-High' with 85% accuracy. Responses were unspeeded and accuracy was encouraged.

### **Seashore pitch test (Seashore, Lewis and Saetveit, 1956)**

The Seashore pitch test was used to test unspeeded pitch discrimination ability. This choice of test represented a change from the test used on the same children in Chapter 5. The change was made because it was considered that there was a degree of ceiling effect on the scores from the Bentley pitch test when the children were aged 11 and that this was likely to become worse when they were aged 13 years. Therefore, the Seashore test, which preserves continuity with the pitch discrimination test used in the study of AIT and the RAZ tests in Chapter 6, was selected. This was administered to the subjects in the manner described in Chapters 5 and 6. Briefly, the test was played on a UHER reel-to-reel tape recorder at high speed. The test was relayed via an auditory network to headphone sets in individual quiet basement cubicles. Subjects were given detailed instructions for the test in small groups and were shown how to complete the response sheet. Further detailed instructions were contained on the test recording. Subjects were required to state the temporal order of two tones which were unmasked and which varied in their pitch difference. The Seashore test uses a method of constant stimuli (descending series) psychophysical procedure.

### **7.2.4 Procedure**

Subjects attended the Department of Psychology for one half day for testing. Subjects attended in groups of about 6 at a time. Subjects first undertook the Mill Hill Vocabulary test in a quiet classroom under examination conditions. After a five minute break, the subjects completed the Raven's Standard Progressive Matrices test in the same room under examination conditions. There then followed a break of about 30 minutes during which subjects were offered fruit juice and a snack. After a description of the Seashore test and instructions on how to complete its response sheet, subjects moved to their individual

basement cubicles, donned their headphone sets and completed the test. Following this, subjects were given a description of the Raz test and were instructed individually on beginning the test. Because each subject was tested individually at their own computer, which controlled stimulus parameters, the tester was able to ensure that all subjects understood what was required of them and that they were responding correctly to the easy discriminations which are met at the beginning of the test. Subjects completed the test when the PEST adaptive staircase procedure had determined the pitch difference at which they were responding with 85% correctness. Therefore, subjects completed this test at different times.

After a break of at least 10 minutes, subjects were given a description of the AIT test, which they had taken about two years before. A detailed description of the test was given, and subjects were instructed on how to complete the response sheet. Subjects then returned to their individual basement cubicles and donned their headphone sets. The AIT test was played on a UHER reel-to-reel tape recorder at high speed and relayed via the auditory network. The test was administered exactly as described in Chapters 4 and 5, with the exception that the stimulus durations of 6ms and 11ms were not used.

### **7.3 Results**

For the purposes of following up some of the findings in the previous study, this section will treat the results of the present study as a straightforward cross-sectional study. Comparisons across the two years of the longitudinal study will be the exclusive concern of Chapter 8.

#### **7.3.1 Results for all subjects (n=108)**

##### **Descriptive statistics and group comparisons**

Table 7.1 shows the means and standard deviations of the measures used in the study for all subjects recalled to the study, and broken down by whether or not each subject was

classified as being able to perform the AIT task. Before discussing the results of particular importance here, it is notable that Table 7.1 indicates that the Raven's Standard Progressive Matrices IQs appear to have increased considerably from the previous testing of these subjects and that the Mill Hill IQs have held (although the Senior as opposed to the Junior form of the test was used for the present study, making this test not identical to the previous test). The other notable descriptive statistic involving all subjects is the Raz test standard deviation, which is large, owing to a number of large estimates. Therefore, for the purposes of correlation logarithmic transformations of the Raz test scores were used.

As can be seen in Table 7.1, subjects classified as being not able to perform the AIT test, using the criteria described in Chapter 4, were no different in age when compared with those who were classified as being able to complete the test. The mean number of items correct in the block of 110 AIT trials (11 durations X 10 trials at each duration) for the group not able to perform the test was 60.1 (SD 7.2). This was close to the level that might be expected by chance (i.e. 55), but did differ significantly from chance levels (one group  $t$  test=4.93,  $p<0.001$ ). When compared with those children able to perform the AIT test, subjects not able to perform the AIT test had lower Raven ( $p<0.01$ ) and Mill Hill ( $p<0.01$ ) IQ scores and scored more poorly on both the Seashore ( $p<0.001$ ) and Raz ( $p<0.001$ ) tasks.

### **Auditory test intercorrelations**

Seashore test scores and logarithmic transformations of Raz test estimates were available for 107 subjects. For the purposes of comparing a measure of performance in the AIT task with other variables in this large number of subjects, a measure of the total number of correctly solved items in the AIT block of 110 items was used. This is referred to as AITcorr or AITcorrect. This was done in order to be able to partial out the effects of pitch discrimination ability from the AIT-ability test correlations in the full group. Because subjects who have superior performances on the AIT task have *higher* AITcorrect scores,

**Table 7.1**

Descriptive statistics (means and SD) for all subjects in the study (n=108), and for those who were defined as being able (n=61) or unable (n=47) to perform the auditory inspection time test. Statistical test results are comparisons of those groups of subjects who could or could not perform the auditory inspection time test.

	All subjects	Could do AIT	Couldn't do AIT	t	p
Age (years)	13.7 (0.38)	13.7 (0.36)	13.8 (0.41)	1.14	ns
Mill Hill Vocabulary IQ	110.3 (8.5)	112.4 (8.6)	107.7 (7.5)	2.97	<0.01
Raven IQ	118.6 (13.2)	122.1 (12.8)	114.3 (12.7)	3.16	<0.01
Auditory inspection time:					
a)Number correct	75.1 (15.9)	86.5 (9.8)	60.1 (7.2)	15.6	<0.001
b)Threshold (ms)	-	78.0 (35.4)	-	-	-
Seashore	38.7 (7.0)	42.4 (4.5)	33.4 (7.0)	8.14	<0.001
Raz test (Hz) <sup>a</sup>	29.5 (40.4)	8.7 (12.5)	57.1 (47.6)	7.26 <sup>b</sup>	<0.001

<sup>a</sup>One subject in the 'Couldn't do AIT' group failed to complete the Raz test, because no threshold could be determined. Therefore, the number of Raz results for this group=46.

<sup>b</sup>Because of the heterogeneity of variance between the two groups on the Raz test, the statistic presented is the Z score from a Mann-Whitney U test.



i.e. they discriminate more items correctly, the expected correlations will be in the direction opposite to that of AIT threshold scores which were used in previous Chapters.

Correlations among the three auditory tasks ranged from 0.60 to 0.75 (all  $p < 0.001$ ), and were all in the expected direction, i.e. those subjects who tended to be superior on one of the tasks tended to be superior on the others (Table 7.2).

### **Correlations between auditory and cognitive ability tests**

Table 7.2 shows that the number of AIT items discriminated correctly and the logarithmic transformations of the Raz test thresholds both correlated at or slightly above 0.4 ( $p < 0.001$ ) with both the Raven and Mill Hill IQ estimates. Seashore pitch test discrimination scores correlated at 0.29 ( $p < 0.01$ ) and 0.34 ( $p < 0.001$ ) with the Mill Hill and Raven IQ scores, respectively. Mill Hill and Raven IQs correlated at 0.39 ( $p < 0.001$ ).

### **Partial correlations between auditory and cognitive ability tests**

Because both AIT and Raz tasks were highly correlated with Seashore test scores, and because Seashore scores were correlated significantly in this group with the verbal and non-verbal IQ tests, the partial correlations between Raz and AIT scores and the IQs were computed, controlling for the effects of pitch discrimination as indexed by the Seashore test. These partial correlations are shown in Table 7.3. The mean change in the four correlations between the auditory tests and the cognitive ability tests was 0.12, i.e. the range of the raw correlations was 0.40 to 0.44 and fell to 0.27 to 0.32. All of the correlations remained significant at the  $p < 0.01$  level. Therefore, pitch discrimination ability appeared to mediate the relationships between Raz and AIT tests and IQs to some extent, but the removal of the effects of Seashore scores did not reduce the correlations to non-significant levels.

**Table 7.2**

Correlations among cognitive ability and auditory test variables for all subjects (n=108, except for correlations involving the Raz test, where n=107).

	AITcorr	logRAZ	Seashore	Mill Hill IQ	Raven IQ
AITcorr	-	-0.75***	0.60***	0.42***	0.41***
logRAZ		-	-0.69***	-0.40***	-0.44***
Seashore			-	0.29**	0.34***
Mill Hill IQ				-	0.39***
Raven IQ					-

\*\* p<0.01; \*\*\* p<0.001

**Table 7.3**

Partial correlations among cognitive ability AIT and Raz test variables for all subjects, controlling for Seashore pitch test score (n=108, except for correlations involving the Raz test, where n=107).

	AITcorr	logRAZ	Mill Hill IQ	Raven IQ
AITcorr	-	-0.59***	0.32***	0.27**
logRAZ		-	-0.29**	-0.30**
Mill Hill IQ			-	0.33***
Raven IQ				-

\*\* p<0.01; \*\*\* p<0.001

### **Factor analysis of auditory and cognitive test scores**

In order to describe more fully the relationships among the three auditory tests, a principal components analysis with Varimax rotation was performed on the scores of the 107 subjects on the Raz, AIT and Seashore tests. The results are shown in Table 7.4. Two components accounted for 92.4% of the variance in the scores. The first unrotated principal component had high loadings for all three of the tasks and accounted for 78.7% of the variance. After Varimax rotation the common variance was distributed between two orthogonal components which accounted for 56.6% and 43.4% of the common variance, respectively. The AIT task and the Raz task had very high and high loadings on the first factor, respectively, whereas the Seashore task had a loading of 0.33. The second factor had a very high ( $>0.9$ ) loading for the Seashore task, a moderate loading for the Raz task and a low loading for the AIT task. Factor scores for each of these two factors were computed for each of the 107 subjects. Table 7.4 shows that Mill Hill IQ correlated at 0.41 ( $p<0.001$ ) with the first factor and at a low level (a near trend) with the second factor. Raven IQs correlated significantly with both factors, but at a higher level with Factor 1.

Because of the relatively clear separation of the AIT and Seashore tasks in the principal components analysis, and, in addition, the ambivalent position of the Raz task, the first and second rotated factors were tentatively named 'Speed' and 'Pitch', respectively. This naming of factors is done primarily as a shorthand to ease further presentation of statistical results and it was not assumed that relatively pure speed and pitch factors have been realised.

The orthogonal Speed and Pitch factor scores for the 107 subjects were used as variables in a further analysis where they were entered into a principal components analysis with Mill Hill and Raven IQ scores. Table 7.5 shows the results of this analysis. Two components accounted for 71.5% of the variance in the results. The first unrotated factor accounted for 46.4% of the variance and had high loadings ( $>0.7$ ) for Mill Hill IQ, Raven

**Table 7.4**

Principal components analysis of auditory tests for all subjects in the study with full data (n=107). Also shown are the correlations between subjects' Mill Hill IQ and Raven IQ scores and their scores on the two factors extracted after orthogonal Varimax rotation.

	First unrotated principal component	Varimax rotation	
		Factor 1	Factor 2
%variance	78.7	56.6	43.4
AIT (no. correct)	0.88	0.92	0.27
logRAZ	-0.92	-0.77	-0.51
Seashore pitch	0.85	0.33	0.93

Correlations between rotated factor scores and ability tests,

	Factor 1	Factor 2
Mill Hill IQ	0.41***	0.16+
Raven IQ	0.39***	0.23*

+ p approx 0.1; \* p<0.05; \*\*\* p<0.001

**Table 7.5**  
Principal components analysis of factors derived from auditory tests after orthogonal Varimax rotation and cognitive ability tests for all subjects with full data (n=107).

	First unrotated principal component	Varimax rotation	
		Factor 1	Factor 2
%variance	46.4	62.2	37.8
Mill Hill IQ	0.77	0.76	0.15
Raven IQ	0.78	0.72	0.31
Factor 1 (Speed)	0.72	0.82	-0.20
Factor 2 (Pitch)	0.36	0.07	0.96

IQ and Speed factor scores. The loading for the Pitch factor was lower, at 0.36, but this level of loading is probably significant beyond the 0.01 significance level (Child, 1990, Table C.2). These results indicate that the Speed factor has a high loading on a factor that also reflects the mental ability that is common to the Raven and Mill Hill IQ tests and that the Pitch factor also has a significant but lower loading on this 'general ability' factor.

Varimax rotation of these components resulted in two factors which accounted for 62.2% and 37.8% of the common variance, respectively (Table 7.5). The first factor had high loadings for Mill Hill IQ, Raven IQ and Speed factor scores and a near-zero loading for the Pitch factor scores. The second factor had a very high ( $>0.9$ ) loading for the Pitch factor, a low but probably significant loading for the Raven IQ scores and non-significant loadings for the Mill Hill and Speed factor scores. Therefore, a factor (Speed) which has a very high loading for AIT scores and a high loading for Raz thresholds was more closely related to the factor that was common to verbal and non-verbal IQ scores than was a factor (Pitch) that had high loadings for the Seashore test and moderate loadings for the Raz test.

### **7.3.2 Subjects who could perform the AIT task (n=61)**

The same analyses as those described above were repeated on those subjects who were classified as being able to perform the AIT test. Table 7.1 shows that this group had significantly higher scores on the Seashore pitch test and that the standard deviation of Seashore scores was somewhat smaller than in the subject sample as a whole. Perusal of the frequency distributions of the Seashore test scores for the groups able and not able to perform the AIT tests revealed that the former group had a slight ceiling effect on the test. Nevertheless, there was still a spread of pitch discrimination ability in the group able to perform the AIT test. It was argued earlier that pitch might act as a threshold factor and that it might have reduced importance as a correlate of other mental abilities beyond a certain levels of pitch discrimination ability. Therefore, following analyses were conducted in order to discover whether this was the case for the group able to perform the



AIT test. Also, the analyses were performed to preserve continuity with previous chapters, where it was assumed that it was prudent to analyse the AIT-cognitive ability test interrelationships in those reckoned to have sufficient pitch discrimination ability to allow a valid estimate of their AIT threshold to be made.

### **Auditory test intercorrelations**

Because the subgroup of subjects examined in this analysis were able to perform the AIT test, both the number of correct discriminations in the AIT block of trials and the AIT thresholds in milliseconds, derived as described in Chapters 4 and 5, were able to be derived. Table 7.6 shows that the correlation between these two estimates of AIT ability were very highly correlated ( $r=0.88$ ,  $p<0.001$ ). The correlations between the two AIT estimates and the Seashore pitch test scores were similar, at 0.24 and -0.23 (both  $p<0.1$ ). When these were corrected for the restriction of Seashore pitch ability range found in this group, the disattenuated correlations were 0.36 and -0.35, respectively, which still falls short of the values found in the subject group as a whole. The correlation between the logarithmic transformation of the Raz test scores and Seashore pitch was -0.39 ( $p<0.01$ ), which became -0.55 after correction for restriction of range on the Seashore test.

### **Correlations between auditory and cognitive ability tests**

Both estimates of AIT ability correlated at just above 0.4 (both  $p<0.001$ ) with Mill Hill IQ scores in this group (Table 7.6). Correlations between AIT estimates and Raven IQ were somewhat lower at 0.24 ( $p<0.1$ ) and -0.30 ( $p<0.05$ ). Raz test scores also correlated at higher levels with the Mill Hill IQ scores than with the Raven IQs. The correlations of the AIT and Raz tests with Mill Hill IQ in this group were similar to those of the whole sample, but the correlations with Raven were reduced in magnitude. The Raven-Mill Hill intercorrelation was 0.39 ( $p<0.01$ ).

**Table 7.6**

Correlations among cognitive ability and auditory test variables for those subjects able to perform the AIT test (n=61).

	AIT(ms)	AITcorr	logRAZ	Seashore	Mill Hill IQ	Raven IQ
AIT(ms)	-	-0.88***	0.47***	-0.23+	-0.41***	-0.30*
AITcorr		-	-0.47***	0.24+	0.42***	0.24+
logRAZ			-	-0.39**	-0.45***	-0.17
Seashore				-	0.18	0.17
Mill Hill IQ					-	0.39**
Raven IQ						-

+ p<0.1; \* p<0.05; \*\* p<0.01; \*\*\* p<0.001

**Table 7.7**

Partial correlations among cognitive ability, AIT and Raz test variables for those subjects able to perform the AIT test, controlling for Seashore pitch test scores (n=61).

	AIT(ms)	AITcorr	logRAZ	Mill Hill IQ	Raven IQ
AIT(ms)	-	-0.87***	0.43***	-0.39**	-0.29*
AITcorr		-	-0.43***	0.39**	0.21+
logRAZ			-	-0.42***	-0.11
Mill Hill IQ				-	0.37**
Raven IQ					-

+ p<0.1; \* p<0.05; \*\* p<0.01; \*\*\* p<0.001

### **Partial correlations between auditory and cognitive ability tests**

Table 7.7 shows the partial correlations among the AIT, Raz and cognitive test scores controlling for the effects of Seashore pitch ability. The change in correlations between the two cognitive ability tests and the Raz and AITcorrect scores was considered, and the mean fall in correlations after pitch was controlled for was 0.03 as opposed to 0.12 for the whole subject sample. The change in the other correlations was similarly small in magnitude.

### **Factor analysis of auditory and cognitive test scores**

Table 7.8 shows the results of a principal components analysis involving the scores on the auditory tests. AITcorrect scores were used here to preserve continuity with the analyses performed on the whole subject sample. Two components accounted for 86.8% of the total test variance. The first unrotated principal component accounted for 63.9% of the variance and had high loadings ( $>0.7$ ) for all three tests. After Varimax rotation two orthogonal factors accounted for 57.4% and 42.6% of the common variance, respectively. The first factor had a very high loading for AIT, a high loading for RAZ and a low (0.19) loading for Seashore pitch. The second factor had a very high loading for Seashore pitch, a moderate loading for Raz (0.41) and a near-zero loading for AIT. Again, the two factors were dubbed 'Speed' and 'Pitch', respectively.

Scores for each of the rotated auditory test factors were calculated for each of the 61 subjects able to perform the AIT test. The correlations between the two cognitive ability tests and the two factors are shown in Table 7.8. Significant correlations were found between Mill Hill and Raven IQ and the Speed factor, with the Mill Hill correlation at 0.49 ( $p<0.001$ ). The correlations with factor 2 were small and non-significant but, nevertheless, they only differed by 0.02 and 0.09 from the respective correlations performed on the whole subject sample.

The scores on the rotated auditory test factors and the Mill Hill and Raven IQs were entered

**Table 7.8**

Principal components analysis of auditory tests for all subjects in the study classified as being able to perform the AIT test (n=61). Also shown are the correlations between subjects' Mill Hill IQ and Raven IQ scores and their scores on the two factors extracted after orthogonal Varimax rotation.

	First unrotated principal component	Varimax rotation	
		Factor 1	Factor 2
%variance	63.9	57.4	42.6
AIT (no. correct)	0.80	0.93	0.08
logRAZ	-0.87	-0.77	-0.41
Seashore pitch	0.72	0.19	0.97

Correlations between rotated factor scores and ability tests,

	Factor 1	Factor 2
Mill Hill IQ	0.49***	0.14
Raven IQ	0.23*	0.14

\*  $p < 0.05$ ; \*\*\*  $p < 0.001$

into a principal components analysis. The results are shown in Figure 7.9. Two components accounted for 70.1% of the score variance. The first unrotated component accounted for 44.8% of the total variance and had high loadings for scores on the two cognitive ability tests and for the Speed factor. The loading for the Pitch factor was 0.27, which is of borderline significance (Child, 1990, Table C.2). Varimax rotation divided the common variance between two factors that accounted for 62.3% and 37.7% of the variance. The first rotated factor had loadings above 0.8 for Mill Hill IQ and Speed and a moderately high loading for Raven IQ. The Pitch loading was near to zero. The second factor had a very high loading for Pitch, and small loadings for the other three variables, though the Raven IQ loading, at 0.32, might be considered significant.

### **7.3.3 Confirmatory factor analyses**

Because confirmatory factor analysis is a method of analysis that is less widely known than most of the other, more standard analyses used in this thesis, the analysis will be described in some detail. The intention of this analysis was to construct models of test relationships that might best explain the variance shared among the auditory and cognitive ability tests. The variables used in the analysis were the putative Speed and Pitch latent variable scores derived from the exploratory factor analyses and the IQ scores from the two cognitive ability tests. Models of the variable relationships were constructed for the whole sample of subjects ( $n=107$ ) and for the subsample of subjects comprising those who were classified as being able to perform the AIT test ( $n=61$ ). The latter group was small for the purposes of confirmatory factor analyses and any results from this group must be treated with caution.

The EQS Structural Equations Program (Bentler, 1989) was used to test the goodness-of-fit of the models. For each group, various competing models were tested and, as is conventional, those models which had the best combination of explanatory parsimony and most acceptable goodness-of-fit will be presented.

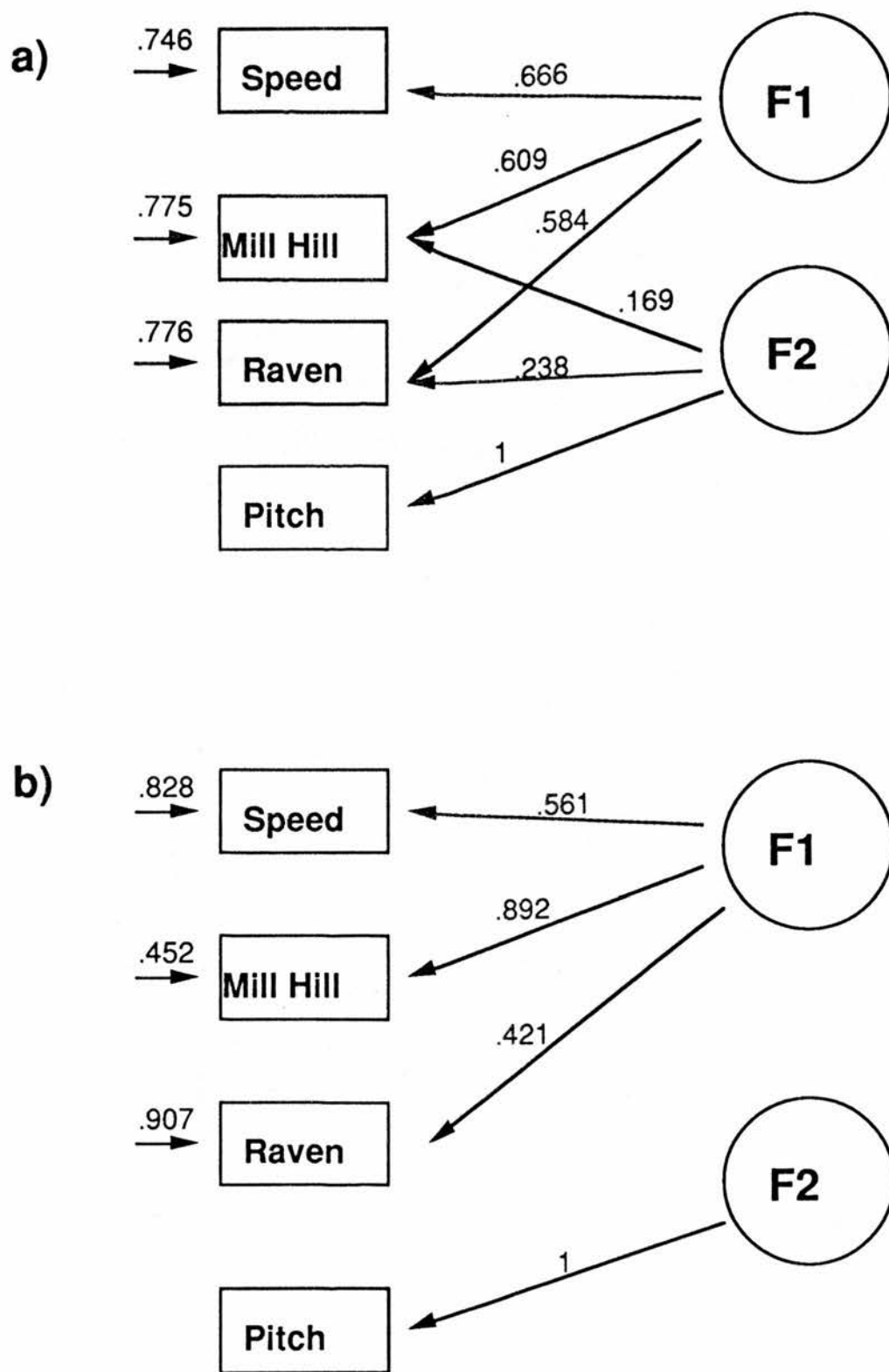
**Table 7.9**  
Principal components analysis of factors derived from auditory tests after orthogonal Varimax rotation and cognitive ability tests for all subjects classified as being able to perform the AIT test (n=61).

	First unrotated principal component	Varimax rotation	
		Factor 1	Factor 2
%variance	44.8	62.3	37.7
Mill Hill IQ	0.85	0.84	0.13
Raven IQ	0.69	0.63	0.32
Factor 1 (Speed)	0.73	0.80	-0.20
Factor 2 (Pitch)	0.27	0.04	0.95



For the whole sample the model in Figure 7.1a was tested. Two latent variables (F1 and F2) were hypothesised to underlie the test interrelationships. Both latent variables were hypothesised to have significant loadings for both cognitive ability tests. However, the intercorrelation between the two latent variables was fixed at zero. The Speed and Pitch factors were derived from Varimax rotation of the scores on the three auditory tests and were, therefore, uncorrelated. Speed and Pitch were allowed to have loadings on only one of the latent variables and the loading on the other latent variable was fixed at zero. All non-zero loadings were free parameters in the model, i.e. their size was allowed to be estimated by the EQS statistical package at levels which gave the best fit to the data. There was one exception; the loading of the Pitch factor on latent variable F2 was set to 1. This may be seen in Figure 7.1a where all the variables except Pitch have error variables as parameters of the model. The covariance matrix of the test scores was used in the analysis. The method of generalised least squares (GLS) was used to assess the adequacy of the model.

After running the EQS analysis on the model in Figure 7.1a the average of the absolute standardised residual correlations among the variables was 0.0122. The chi square for the model was 0.344 (2 d.f.,  $p=0.84$ ). Higher  $p$  values indicate better fitting of models to the data, and significant levels of  $p$  are often used to indicate that the model has a poor fit. In the present case, the model appeared to fit very well. However, the chi square must be assessed in relation to the number of subjects on which the model was based. The present  $n$  of 107 is not high enough to allow great confidence to be placed in the chi square alone; a much larger number of subjects might have led to the model having a significant  $p$  value. Other indications of model acceptability include the fit indices provided by EQS. The Bentler-Bonett Normed fit index for the model was 0.999. Values greater than 0.9 are usually taken to indicate acceptable fit of the model to the data (the maximum value for this statistic is 1). The Bentler-Bonett Non-normed fit index, which is sensitive to the degrees of freedom in the model, was 1.013 (values slightly in excess of 1 are not uncommon for



**Figure 7.1**

Path diagrams obtained from the EQS package for models of the interrelationships among the factors derived from the auditory tests and the cognitive ability tests for: a) all subjects ( $n=107$ ); and b) those subjects able to perform the AIT task ( $n=61$ ). F1 and F2 are latent factors assumed to underlie the various test scores. Numbers are parameter estimates of the models. Arrows without origins represent task-specific and error variance.

well-fitting models with this index). Therefore, the model appears to fit well by the various criteria used to assess goodness-of-fit.

The standardised solution of the model may be summarised as a series of measurement equations expressing the relationships among the variables. In the present models each measured variable is expressed in terms of latent variables and variables representing residual and error terms, except in the case of pitch which had no residual/error component. The parameters of these equations are shown as the numbers adjacent to the arrows in Figure 7.1a. These represent the optimal parameter estimates as determined by the EQS program. Parameters may be tested to discover whether their size differs significantly from zero, i.e. whether certain of the paths might be omitted from the model without reducing the fit of the model. This is done by dividing the parameter estimates by their standard errors. This gives a test statistic which, in the context of an appropriate model, is a univariate large-sample normal z-test of the null hypothesis that the parameter in question is zero in the population. All of the free parameters in the model (Figure 7.1a) were significant beyond the 0.05 level (i.e. they had test statistics greater than 1.96). Therefore, all of the relationships expressed in the model contribute significantly to its success.

The sum of the squares of the parameters leading to each variable in the standardised model is 1, i.e. all the variance in the measured variable is accounted for in terms of other variables (in this case latent variables) or by residual/error variables. For instance, from a perusal of Figure 7.1a, it may be calculated that 44.4% of the variance in the Speed factor may be accounted for by a latent variable that underlies Mill Hill and Raven performance in addition to Speed factor scores. This also allows us to conclude that 55.6% of the Speed factor variance is unaccounted for by the latent variable. It may also be calculated that about 6% of Raven IQ variance and 3% of Mill Hill IQ variance is accounted for by the latent variable on which the Pitch factor has a loading of 1.

The EQS program offers an alternative to the testing of several models consecutively. Implementation of the Wald test informs the user whether any of the free parameters in the model might be considered zero without a significant reduction in the goodness-of-fit of the model (Lee, 1985; Satorra, 1989; Bentler, 1989). The Wald test was used in the present case and led to none of the parameters being dropped, i.e. all the free parameters in Figure 7.1a were necessary for optimal model fit. Further, the Lagrange Multiplier test was applied (Bentler, 1989). This assesses whether parameters that have been fixed at certain values might be freed in order to improve the fit of the model. In the present case the results of the test implied that there would be no improvement in model fit from the freeing of any fixed parameters.

The same procedure as has been described for the assessment of fit of the model in Figure 7.1a was undertaken for the data collected from the 61 subjects classed as being able to perform the AIT test. Again, caution must be urged owing to the limited sample size for this type of analysis. The model shown in Figure 7.1b proved to have the best fit to the data, i.e. a model with two uncorrelated latent variables, one of which has a parameter value fixed at 1 for the path connecting it to the Pitch factor. The mean of the absolute standardised residuals for this model was 0.0459. The chi square for the model was 2.024 (2 d.f,  $p=0.36$ ). The Bentler-Bonett Normed fit index was 0.991 and the Bentler-Bonett Nonnormed fit index was 1.000. Therefore, the model had acceptable fit indices. The Wald and Lagrange Multiplier test results indicated that there were no free parameters that might be fixed and no fixed parameters that might be freed in order to improve the fit of the model.

Factor 1 in Figure 7.1b represents a latent variable underlying performance on the verbal and non-verbal ability tests and on the Speed factor. About 31.5% of the Speed factor variance is accounted for by this factor. It is notable that, in this model, latent variable F1

is heavily biased toward verbal IQ; about 80% of the Mill Hill IQ variance is accounted for by this factor, whereas only about 18% of the Raven IQ variance is accounted for by F1. This reflects the closer correlation between AIT and Mill Hill IQ than between AIT and Raven IQ in this subgroup (Table 7.6).

Therefore, with the reservations that have been expressed above, it may be concluded that the Pitch factor was significantly associated with verbal and non-verbal IQ in the whole sample, but not in the subjects classified as being able to perform the AIT test, among whom only the Speed factor was associated with individual differences in cognitive ability.

#### **7.4 Discussion**

The results of the above study confirm the impression of Chapter 6 that all three auditory tests - AIT, Raz and Seashore - were significantly intercorrelated. Principal components analysis of the auditory tasks revealed a general auditory ability test factor, and Varimax rotation supported the hypothesis that AIT and Seashore proved to be relatively separable, with the Raz task at least moderately highly loaded on both tasks. It appeared to be reasonable to call the rotated factor with a very high loading on Seashore, a pitch discrimination factor. The factor with high loadings for AIT and Raz was called a Speed factor, congruent with the notion that AIT assesses speed of auditory processing. However, alternative interpretations of this factor are possible and it is prudent to consider these. For instance the 'Speed' factor might represent a general task complexity or difficulty factor, reflecting the fact that the brief nature of the stimuli in the Raz and AIT tasks makes the performance of the task generally more difficult than the Seashore task.

Individual differences on all three auditory tests were related significantly to scores on the tests of cognitive ability. This is the first study in this thesis in which a large number of subjects was tested on the Seashore test, and there were significant associations between Seashore pitch scores and the verbal and non-verbal ability test scores for the whole group,

though these fell to non-significant levels when only those who were classified as being able to perform the AIT task were considered. This latter result probably reflects the fact that the group classified as being able to perform the AIT task were essentially being selected for their relatively good pitch discrimination skills. For the group as a whole the auditory tests appeared to be associated equally closely with the verbal and non-verbal tests, but there was a closer association between the AIT and Raz tests and verbal ability in the subgroup classified as being able to perform the AIT test.

In support of the stronger association with cognitive ability that held between the Raz and AIT tests as opposed to the Seashore test, confirmatory factor analysis showed that the so-called Speed factor was clearly more highly related to a factor that was general to Raven and Mill Hill IQs than was the Pitch factor. The latter factor was significantly associated with a factor that was general to Mill Hill and Raven IQs in the whole group, but not in the group classified as being able to perform the AIT task. Again, this probably reflected the attenuation of pitch discrimination ability in the latter group and, perhaps, the fact that there is a threshold of pitch discrimination above which individual differences in pitch discrimination are no longer correlated with individual differences in cognitive ability test scores.

If Figure 7.1a is considered it appears that variance shared by Mill Hill and Raven IQs was significantly associated with two orthogonal factors, Speed and Pitch. A similar result, using factors derived from reaction time scores, was taken by Vernon, Weese and Miller (1991) to indicate that 'g' was not unitary and that it consisted of separate, independent basic abilities rather than some single source of variance common to all of the basic abilities underlying intellectual performance. However, the conclusion of Vernon, Weese and Miller (1991) is not necessarily correct. The separate factors to which the IQs are related do not represent 'things' that are necessarily real. They are factors derived from various auditory tests, and it is quite possible that the loading of the cognitive ability tests with both

of these factors comes about because the 'real' basic ability related to 'g' is captured by a vector somewhere between the two orthogonal factors, i.e. it might be more closely represented by the first unrotated factor on which all three auditory tests had high loadings. The main purpose of the present study was to discover the degree to which speed and pitch discrimination proved separable, given the tests of auditory processing that were of interest. However, more emphases might have been placed on the variance shared by auditory tests; in fact, when scores on the first unrotated principal component extracted from the three auditory tests were correlated with the scores on the first unrotated component extracted from the Mill Hill and Raven IQs, the Pearson's  $r$  was 0.52 ( $p < 0.001$ ;  $n = 107$ ).

In summary, auditory tests which were designed to measure separate pitch discrimination and speed of processing abilities are intercorrelated. However, it was argued that these proved separable to some extent and that, when this was done, speed of processing was more closely related to cognitive ability test scores, though pitch discrimination had a significant association with cognitive ability in the whole sample. Exploratory factor analyses helped to extend and clarify the correlational analysis, and confirmatory factor analysis was used to test models of the relationships among auditory processing and cognitive ability.



## **Chapter Eight**

### **Auditory inspection time and cognitive ability: a longitudinal study**

#### **8.1 Introduction**

The present Chapter will present the results of a longitudinal study of AIT ability conducted over two years. This was made possible by recalling the schoolchildren who provided data for Study 2 in Chapter 5. Nettelbeck and Wilson (1985, Study 3) showed that improvements in IT performance, tested twice in the same group of subjects over a year, could not be explained entirely in terms of the effects of practice or task-specific knowledge. They found that improvements in a second cohort of 6- and 11-year-old children and university student adults tested on IT twice over a two week period were much smaller than the improvements in the original cohort over a year. Further testing at a second year of follow up continued to indicate developmental differences in IT performance up to, and perhaps beyond, age 13, and there were no cohort effects, which are a potential problem for longitudinal studies (Baltes, 1968; Sugarman, 1986).

Arguably, the most important issue concerning the IT-cognitive ability association is the question of the direction of causation. If IT performance is simply another high-level cognitive task or learned skill that high IQ individuals find ways to perform better, then IT may be added to the large number of higher cognitive aspects of human performance that are predicted by mental test scores (Brand, 1987a). Another possibility is that IT performance might be one of the basic processing skills upon which intellectual differences are based (see Chapter 1). The present study attempts to address the issue of causality between AIT and cognitive ability test scores by using a cross-lagged panel design and then subjecting the results to formal testing by structural equation modelling, as was done by Bentler and Speckart (1981) in a different area of individual differences research in their

landmark "Attitudes 'cause' behaviour" paper. A similar design was used by Ecob (1987) to demonstrate that learning difficulties in children can lead to later reading problems.

The cross-lagged panel design has already been used by Nettelbeck and Young (1990) in a study of visual IT and cognitive ability in children. Nettelbeck and Young (1990) provided a useful summary of the thinking behind the cross-lagged panel design,

...the logic of this design permits inference about a causal relationship on the basis of correlation, because correlations are crossed and lagged over time. Specifically, if IT causes IQ, then the correlation between IT on the first occasion ( $IT_1$ ) and IQ on the second ( $IQ_2$ ) should be reliably greater than the  $IQ_1$ - $IT_2$  correlation. If the reverse held, then IQ would be inferred to cause IT performance. However, the third possibility, with cross-lagged correlations not significantly different, would indicate no causal relationship, although both outcomes could be caused by some common influence.

Nettelbeck and Young (1990) retested thirty 7-year-old children from an original sample of 47 who were tested from 11 to 15 months earlier (Nettelbeck and Young, 1989), on visual IT and on the Wechsler Intelligence Scale for Children-Revised. Reliability of visual IT across this period was 0.40 ( $p < 0.05$ ). The cross lagged correlations were both -0.40, and Nettelbeck and Young (1990) concluded that the relationships between the variables was not causal, though IT and IQ, they hypothesised, might both be caused by a general ability factor, as suggested by Mackintosh (1986). Thus, they found no support for the suggestion by Brand (1981, 1984) that IT might be causally related to IQ differences.

There was no requirement for Nettelbeck and Young to carry out comparisons of the cross-lagged correlations in their study because they were identical. In the present study, a similar design to that used by Nettelbeck and Young (1990) was used to test hypotheses concerning the causal relation between AIT and cognitive ability. In addition, the following methodological improvements were made: the AIT test was identical on the two occasions and was tested in the same conditions, the subject numbers were larger, and formal models of causal relationships were tested using goodness-of-fit statistics (Bentler,

1989).

## **8.2 Method**

### **8.2.1 Subjects**

The subjects who provided data to be used in the analyses were those schoolchildren who were able to attend the Department of Psychology on two occasions, two years apart.

They are the same subjects provided data for the studies described in Chapters 5 and 7.

Their age and sex characteristics were described in the earlier chapters.

### **8.2.2 Cognitive ability tests**

#### **Verbal ability testing**

The Mill Hill Vocabulary Form 1 Junior (Synonyms and Definitions) was administered during their first visit and the Mill Hill Vocabulary Form 1 Senior (Synonyms and Definitions) was administered during their second visit, two years later, as described in the Method sections of Chapters 5 and 7, respectively.

#### **Non-verbal ability testing**

The Raven's Standard Progressive matrices test was administered on both occasions, as described in the Method sections of Chapters 5 and 7.

### **8.2.3 Auditory inspection time**

The same auditory inspection time test was administered on both testing occasions, with the exception that the 11ms and 6ms stimulus durations were omitted from the test on the second occasion. The details of the test stimuli, psychophysical procedure and the testing and scoring procedures are fully described in Chapters 4, 5 and 7, respectively. Briefly, a method of constant stimuli (descending series) was used to find the briefest stimulus duration at which the subjects were 90% accurate in discriminating a 96 Hz difference

between two consecutively played 80 dB tones which were followed by an auditory mask. In order to maximise the number of subjects included in the longitudinal study, the number of correctly discriminated items from the block of 110 items was taken as an additional measure of the AIT ability of subjects. This allowed an AIT variable (AITcorr or AITcorrect) to be derived for all subjects whereas the AIT threshold (AIT(ms)) could only be derived from the data provided by those subjects classified as being able to perform the AIT task according to criteria described earlier.

### **8.3 Results**

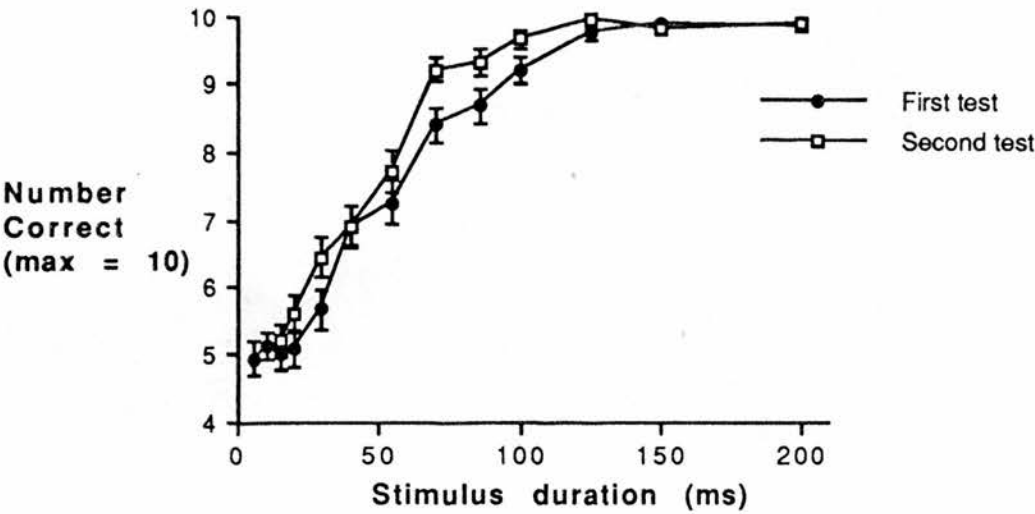
#### **8.3.1 Comparisons across time in cognitive ability and inspection time**

##### **Auditory inspection time**

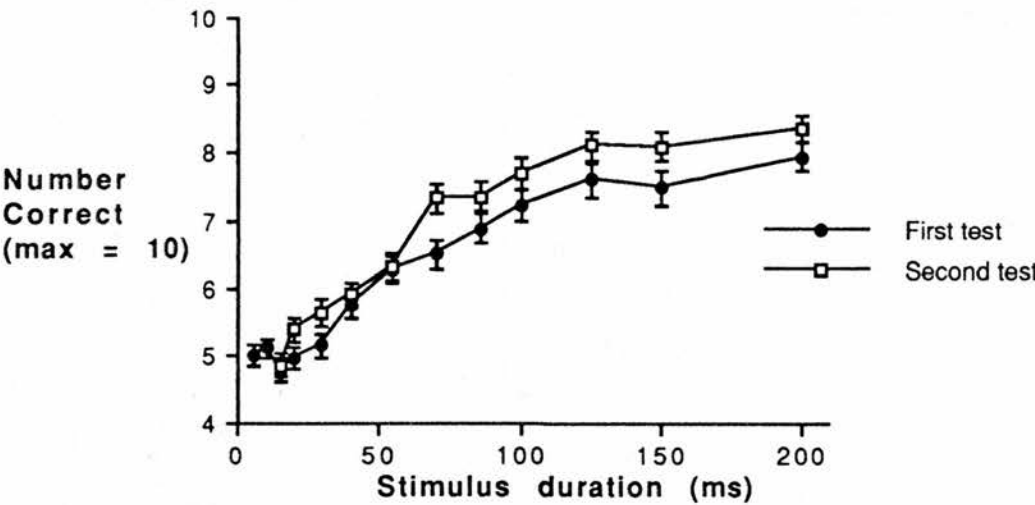
Figures 8.1a to 8.1c show the group psychometric curves for those subjects classified as being able to perform the AIT task on both test visits ( $n=45$ ), for all subjects with full AIT data on both test visits whether or not they were classified as being able to perform the AIT task ( $n=108$ ), and for those subjects classified as not being able to perform the AIT test on both test visits ( $n=47$ ), respectively.

For subjects classified as being able to perform the AIT task on both occasions ( $n=45$ ) there appeared to be an improvement in AIT performance, especially over the stimulus durations 70ms to 100ms, inclusive. The data from these subjects is presented first because they represent a group who were unequivocally able to perform the AIT task. A two way analysis of variance test was performed on these data with Test session (having two levels, i.e. first or second) and stimulus duration (having 11 levels, i.e. 15ms to 200ms) as within subjects factors. The effect of test session was significant ( $F=16.803$ ,  $d.f.=1,44$ ,  $p<0.001$ ), indicating that subjects had improved on the test from the first to the second test session. The effect of duration was significant ( $F=143.065$ ,  $d.f.=10,440$ ,  $p<0.001$ ), i.e. longer stimulus durations led to superior scores. The interaction between

a)



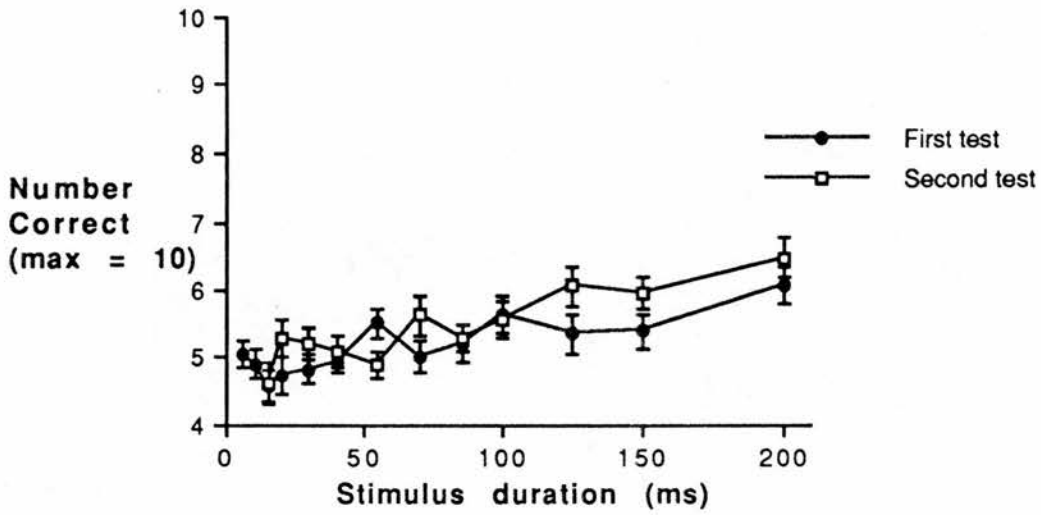
b)



**Figures 8.1a and 8.1b**

Means with standard error bars for the number of correct AIT item responses at different stimulus durations for schoolchildren on two attempts at the AIT task, two years apart. Subjects included in Figure 1a were those classified as being able to perform the AIT task ( $n=45$ ) on both test visits. Subjects included in Figure 1b were all subjects who had AIT data for both visits and includes subjects who could and who could not perform the AIT task ( $n=108$ ). Shaded circle symbols represent the scores on the first test and open square symbols represent the scores achieved when the same test was given two years later.

c)



**Figure 8.1c**

Means with standard error bars for the number of correct AIT item responses at different stimulus durations for schoolchildren on two attempts at the AIT task, two years apart. Subjects included in this figure were those classified as not being able to perform the AIT task on both test visits ( $n=47$ ). Shaded circle symbols represent the scores on the first test and open square symbols represent the scores achieved when the same test was given two years later.

the two scores was not significant ( $F=1.684$ ,  $d.f.=10,440$ ,  $p=0.08$ ), but there was a trend indicating that the test session improved scores on some stimulus durations in preference to others. Simple effects of test session were computed for the different durations. Because of the number of effects computed, the interpretation must be done with caution. The second test session resulted in significantly better performance at the following durations: 125ms ( $p=0.018$ ), 100ms ( $p=0.009$ ), 85ms ( $p=0.001$ ), 70ms ( $p=0.004$ ), and 30ms ( $p=0.039$ ). The simple effects test results support the impression that the most significant improvements over two years were, in the main, at durations above 70ms. The durations above 125ms resulted in scores so near to the asymptote for his group that there was unlikely to be any improvement.

Figure 8.1b contains the results for all subjects ( $n=108$ ) who attended twice across two years for AIT testing. Perusal of the psychometric curves indicates that most benefit from the gap of two years accrues to durations longer than 55ms. Two way ANOVA testing was performed as for the subjects who could do the AIT test. The effect of session was significant, with scores on the second session being better than those on the first ( $F=26.717$ ,  $d.f.=1,107$ ,  $p<0.001$ ). The effect of time was significant, with longer stimulus durations having higher scores than the shorter durations ( $F=81.767$ ,  $d.f.=10,1070$ ,  $p<0.001$ ). The interaction between test session and stimulus duration was not significant ( $F=1.373$ ,  $d.f.=10,1070$ , ns). Simple effects were computed as above and must be treated with similar caution owing to multiple tests being carried out. Significant improvements were found at the second test session in scores for the following stimulus durations: 200ms ( $p=0.022$ ), 150ms ( $p=0.002$ ), 125ms ( $p=0.009$ ), 100ms ( $p=0.024$ ), 85ms ( $p=0.009$ ), 70ms ( $p=0.000$ ), 30ms ( $p=0.025$ ).

Figure 8.1c shows the psychometric curves for the two testing sessions for those subjects ( $n=47$ ) who were, on both occasions, classified as not being able to perform the AIT task. When compared with Figures 8.1a and 8.1b there was much less obvious improvement



across the two years, except at the longest durations. Two way ANOVA testing as described above revealed the overall effect of test session did not reach conventional levels of significance, but there was a trend in the predicted direction ( $F=3.500$ ,  $d.f.=1,46$ ,  $p=0.068$ ). The effect of stimulus duration was significant ( $F=6.906$ ,  $d.f.=10,460$ ,  $p<0.001$ ), i.e. longer stimulus durations resulted in higher scores than did briefer stimulus durations on the whole. The interaction between test session and stimulus duration was non-significant ( $F=1.338$ ,  $d.f.=10,460$ , ns). Because the overall effect of test session only tended toward significance, examination of simple effects was not strongly warranted. However, perusal of simple effects indicated that, in this subgroup, the only stimulus duration which improved significantly over the period of two years was the 55ms duration ( $p=0.051$ ).

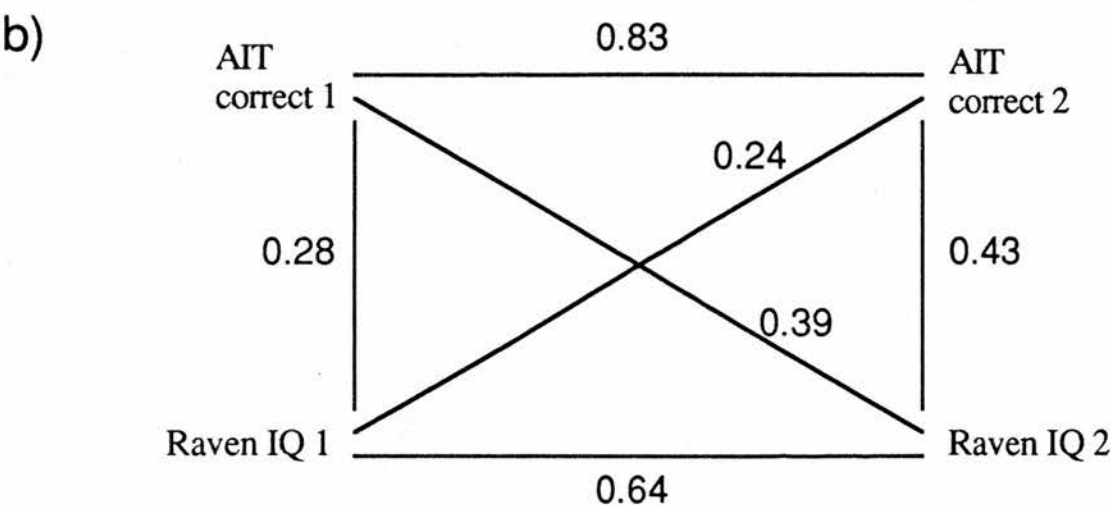
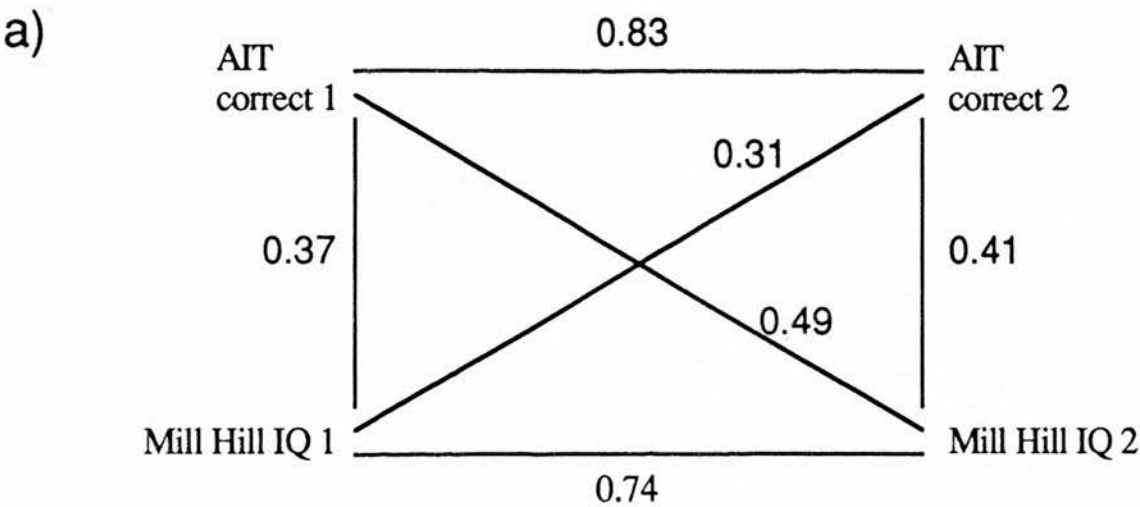
### **Cognitive ability**

Significant improvements occurred across the two year period in the number of Mill Hill Vocabulary items solved; 38.2 (SD 5.8) were solved correctly in the first session versus 44.8 (SD 5.7) in the second session two years later ( $t=16.8$ ,  $p<0.001$ ). A significant increase in the mean number of Raven's Progressive Matrices items solved correctly also occurred over this period; 46.2 (SD 6.1) were solved correctly in the first session versus 50.7 (SD 5.0) in the second session ( $t=9.9$ ,  $p<0.001$ ).

### **8.3.2 Correlations across time for cognitive abilities and inspection time**

#### **All subjects**

Figures 8.2a and 8.2b show the cross-lagged panel correlations for the number of AIT test items solved correctly at test session 1 and test session 2 versus Mill Hill IQ and Raven IQ separately. All correlations involved the same 104 subjects who provided full cognitive ability test and AIT test data in the two sessions. AITcorrect scores were reliable across the two year period, with a correlation of 0.83 ( $p<0.001$ ) between the two test results. The reliability of the Raven's Progressive Matrices IQ was 0.64 ( $p<0.001$ ) and of the Mill



**Figures 8.2a and 8.2b**  
Cross-lagged panels of correlations involving: a) Mill Hill Vocabulary IQ and the number of AIT items discriminated correctly in sessions 1 and 2 ( $n=104$ ), and b) IQ scores from Raven's Progressive Matrices and the number of AIT items discriminated correctly in sessions 1 and 2 ( $n=104$ ).

Hill IQ was 0.74 ( $p < 0.001$ ). The correlations between the cognitive ability tests and the AIT correct scores for the second testing session were presented in Chapter 7. Slight differences in correlations are due to this Chapter's using only those subjects who attended twice and provided full data. The correlations between AITcorrect scores and Mill Hill IQ and Raven IQ for the first testing session were 0.37 ( $p < 0.001$ ) and 0.28 ( $p < 0.01$ ), respectively.

The results of primary interest in Figures 8.2a and 8.2b were the cross-lagged correlations between AITcorrect tests and cognitive ability measures. The difference between the two diagonal correlations is conventionally used to indicate the direction of causation, i.e. if one diagonal correlation is much larger than the other then, for the diagonal arm of the larger correlation, the individual differences in the measure at time 1 are causally implicated in the individual differences in the other measure at time 2 (Crano, Kenny and Campbell, 1972). This section will examine these correlations in three ways, increasing in rigour. First, the differential magnitude of the correlations will be presented. Second, the two diagonal correlations in each panel will be compared to discover whether they differ significantly. Third, a structural modelling exercise will be undertaken to test hypotheses concerning the direction of causation of individual differences in cognitive ability and auditory inspection time, similar to that used by Bentler and Speckart (1981) to study the causal associations between attitudes and behaviour.

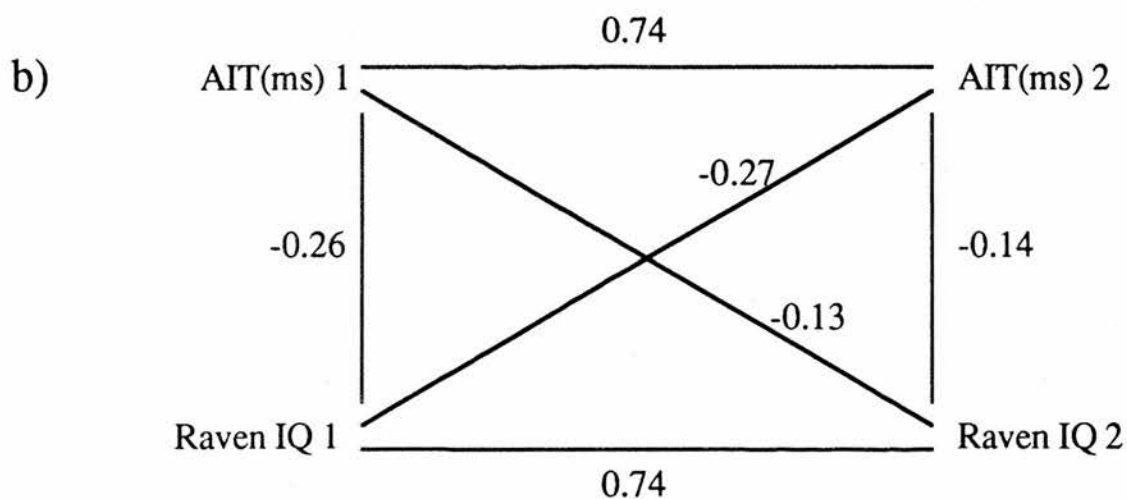
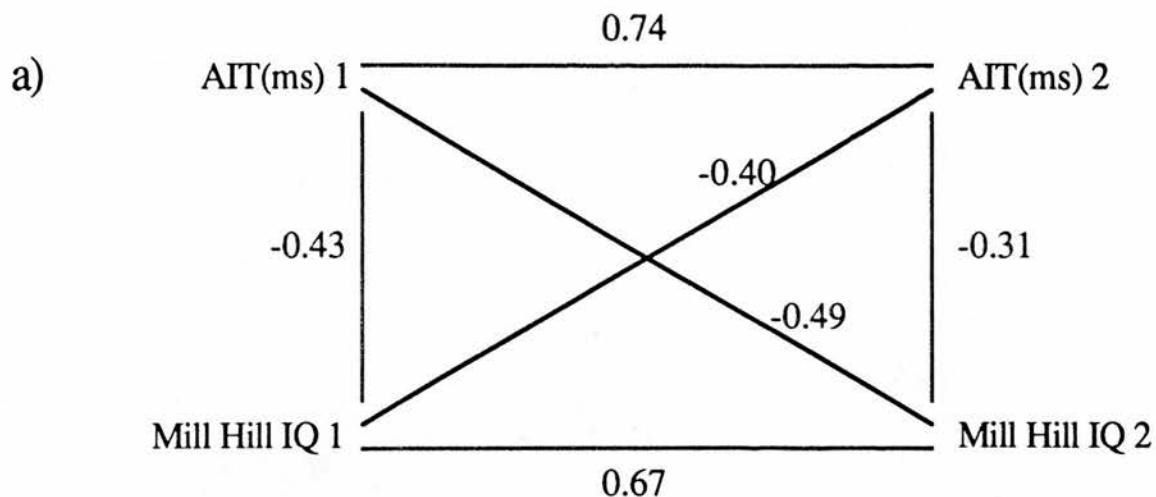
In Figure 8.2a it can be seen that the correlation between AITcorrect at time 1 and Mill Hill IQ at time 2 was 0.49 ( $p < 0.001$ ). The correlation between Mill Hill IQ at time 1 and AITcorrect at time 2 was 0.31 ( $p < 0.01$ ). Therefore, the stronger link lay between AITcorrect scores at 11 years of age and verbal IQ at 13 years than between the earlier verbal IQ and the later AITcorrect scores. Figure 8.2b shows that the correlation between AITcorrect at time 1 and Raven IQ at time 2 was 0.39 ( $p < 0.001$ ), whereas the correlation between Raven IQ at time 1 and AIT correct at time 2 was 0.24 ( $p < 0.05$ ). Therefore, the

stronger association lay in the same direction as was discovered for verbal IQ, i.e. from the earlier AIT to the later IQ.

To discover whether the diagonal correlations were significantly different in the cross-lagged panel in Figure 8.2a, a  $\bar{Z}_2^*$  statistic was used to compare the sizes of the two correlations (Steiger, 1980). Steiger (1980) has argued that this represents the best statistic for testing the hypothesis under study in the present situation, whereas he states that other statistics frequently used to assess the significance of the difference between two dependent correlations are suboptimal or "basically useless". The formula for this hand-calculated statistic is given in Appendix 4. From the correlations presented in Figure 8.2a it was found that the  $\bar{Z}_2^*$  value was 2.19 ( $p < 0.05$ ), indicating that the correlation between AITcorrect on the first occasion and Mill Hill IQ on the second occasion was significantly greater than the correlation between Mill Hill IQ in session 1 and AITcorrect in session 2. Examination of the correlations in Figure 8.2b resulted in a  $\bar{Z}_2^*$  value of 1.37 (ns), indicating that there was not a significant difference between the two diagonal correlations involving Raven IQ and AITcorrect scores measured on two occasions two years apart.

### **Subjects able to perform the AIT task**

Figures 8.3a and 8.3b show the cross lagged panels representing the correlations between AIT thresholds (in ms) and verbal and non-verbal cognitive abilities for those subjects who were classified as being able to perform the AIT task on both test sessions and who had full AIT and cognitive ability data ( $n=43$ ). Reliability correlations across two years for Mill Hill IQ, Raven IQ and AIT thresholds in this group were 0.67, 0.74 and 0.74, respectively. Correlations between AIT thresholds and Mill Hill IQ were all significant (Figure 8.3a). In particular, the cross-lagged correlations were highly significant, with AIT threshold at time 1 correlating with Mill Hill IQ at time 2 at -0.49 ( $p < 0.001$ ) and Mill



**Figures 8.3a and 8.3b**

Cross-lagged correlations for AIT thresholds and a) Mill Hill IQ and b) Raven IQ for all subjects classified as being able to perform the AIT task ( $n=43$ ). Correlations  $>0.26$  are tend toward significance at  $p<0.1$ , those  $>0.30$  are significant at  $p<0.05$ , and those  $>0.39$  are significant at  $p<0.01$ .

Hill IQ at time 1 correlating with AIT threshold at time 2 at -0.40 ( $p < 0.01$ ). Because of the relatively low number of subjects and the small difference between these two correlations, there was little point in formally calculating the  $Z_2^*$  statistic or in performing an analysis of causal models. There was no evidence in this subgroup to indicate that there was a significant difference in these two cross-lagged correlations. Figure 8.3b shows that the four correlation coefficients between AIT thresholds and Raven IQs range from -0.13 to -0.27. Two of the correlations tended toward significance and the other two were non-significant, though all were in the expected direction. The correlation between Raven IQ at time 1 and AIT threshold at time 2 was -0.27, and represented the highest AIT-Raven IQ correlation in this subgroup. Again, it was not necessary to carry out formal comparisons of the cross-lagged correlations, because of the small numbers of subjects and because of the small differences between the sizes of the correlations. Although causal models were not tested it should be noted that, in this subgroup, the relative sizes of the cross-lagged correlations did not harmonise with the structural model results of the Raven IQ-AITcorrect variance for the whole group. The analyses for those subjects who could do the AIT task were also performed using AITcorrect scores. The correlations were very similar to those found with threshold values and, therefore, afforded the same conclusions.

### **8.3.3 Structural modelling of putative causal relationships across time**

The EQS Structural Equations Program (Bentler, 1989) was used to construct and to test putative causal models of the relationships between Mill Hill IQ and AITcorrect scores and between Raven IQ and AITcorrect scores. The traditional cross-lagged panel design of the present study was used to construct multivariate regression models of the relationships between cognitive ability and AITcorrect scores. Test scores on the cognitive tests and on the AIT test taken at time 1 were used as independent variables in the model. If the relationships between the variables were to be constructed assuming a pathway to represent every relationship displayed in the cross-lagged panel shown in Figures 8.2a and 8.2b, the model would be saturated, i.e. there would be as many parameters as there were data

points and, as a result, the model would have zero degrees of freedom. Therefore, for the purposes of hypothesis testing, restrictions must be placed upon the models. In the present case this was done by assuming that certain pathways in the model might be fixed at zero without a significant lowering of goodness-of-fit.

In all of the models tested there was no path indicated between the AIT and cognitive test variables at the second testing session, i.e. it was assumed that the covariation between these variables at the second testing session was accounted for by the covariation between them at the first test session. Three models were tested for best fit to the data in the present study. The same exercise was repeated for the Mill Hill IQ and the Raven IQ data. The first model (the 'Reciprocal causation' model) assumed that AITcorrect at time 1 had a significant influence on cognitive ability at time 2 *and* that Mill Hill IQ at time 1 significantly influenced AITcorrect scores at time 2. Therefore, this model had one degree of freedom, and the only path which has been omitted from the full cross-lagged panel (see figure 8.2) was the AITcorrect 2 versus cognitive ability at time 2 path. Other pathways were assumed to express causal relations, except AITcorrect 1 versus cognitive ability at time 1, where the path expressed a correlation between these independent variables.

The second model tested for each cognitive ability introduced a second degree of freedom. This model (the 'AIT causes IQ' model) omitted the path between cognitive ability (Mill Hill IQ or Raven IQ) at time 1 and AITcorrect scores at time 2, i.e. this path parameter was fixed at zero. This model assumed that AIT scores at time 1 affected cognitive ability at a later point more strongly than cognitive ability at time 1 affected AIT correct scores at a later point. The third model tested (the 'IQ causes AIT' model) also had two degrees of freedom. It omitted the path between AITcorrect scores at time 1 and cognitive ability (Mill Hill IQ or Raven IQ) at time 2. This model assumed that cognitive ability differences at time 1 affected AITcorrect scores at time 2 more strongly than AIT ability at time 1 affected cognitive ability in the later testing session.



### **Models involving Mill Hill IQ and AITcorrect scores**

Covariance matrices expressing the relationships among Mill Hill IQ at times 1 and 2 and AIT correct at times 1 and 2 were used in the analyses. The generalised least squares and the maximum likelihood solutions were computed for the models. Table 8.1 shows the results of the three models for both methods of analysis. For Model 1 ('Reciprocal causation') the chi square values were very low and the p values very high, indicating that the model had a good fit to the data. The fit indices had values of one or slightly greater, indicating acceptable fit also. However, one of the parameters of the model (the path between Mill Hill IQ at time 1 and AITcorrect scores at time 2) had a very low z statistic, indicating a non-significant contribution to the model. In addition, the Wald test indicated that this parameter may have been dropped from the model without a significant worsening of its fit to the data.

Model 2 ('AIT "causes" Mill Hill IQ') also resulted in very low chi square values and high p values, indicating acceptable fit to the data (Table 8.1). Fit indices were at or slightly greater than one. No parameters in this model had non-significant z statistics, indicating that all paths in the model made a significant contribution to its fit. Therefore, this model was more parsimonious than Model 1, it had good fit indices and a non-significant p value and there were no parameters that required to be added to or dropped from the model. The results of the generalised least squares and maximum likelihood solutions for model 3 ('Mill Hill IQ "causes" AIT') are shown in Table 8.1. Both methods of solution resulted in large and highly significant chi square values, indicating a poor fit to the data. The fit indices for the generalised least squares solution were high ( $>0.9$ ), but the maximum likelihood solution Non-normed fit index was lower than 0.9, indicating that the model was unacceptable. In addition, the path linking Mill Hill IQ at time 1 with AITcorrect scores at time 2 did not have a significant z score.

In order to choose the best model of the three from the above results, the criteria of

**Table 8.1**  
 Summary of structural equations program output for models of the relationships between AITcorrect and Mill Hill IQ tested two years apart on the same subjects (n=104).

	Model 1: Reciprocal causation	Model 2: AIT 'causes' Mill Hill IQ	Model 3: Mill Hill IQ 'causes' AIT
<b>(i) Generalised least squares solution</b>			
Chi square (d.f.)	0.033 (1)	0.034 (2)	10.676 (2)
p value	0.855	0.983	0.005
Bentler-Bonett Normed fit index	1.000	1.000	0.996
Bentler-Bonett Non-normed fit index	1.002	1.002	0.991
Non-significant parameter values in the model	Mill Hill 1 vs AIT 2	None	Mill Hill 1 vs AIT 2
Z score of non- significant parameters	-0.012	-	0.037
<b>(i) Maximum likelihood solution</b>			
Chi square (d.f.)	0.033 (1)	0.034 (2)	12.658 (2)
p value	0.855	0.983	0.002
Bentler-Bonett Normed fit index	1.000	1.000	0.945
Bentler-Bonett Non-normed fit index	1.002	1.026	0.858

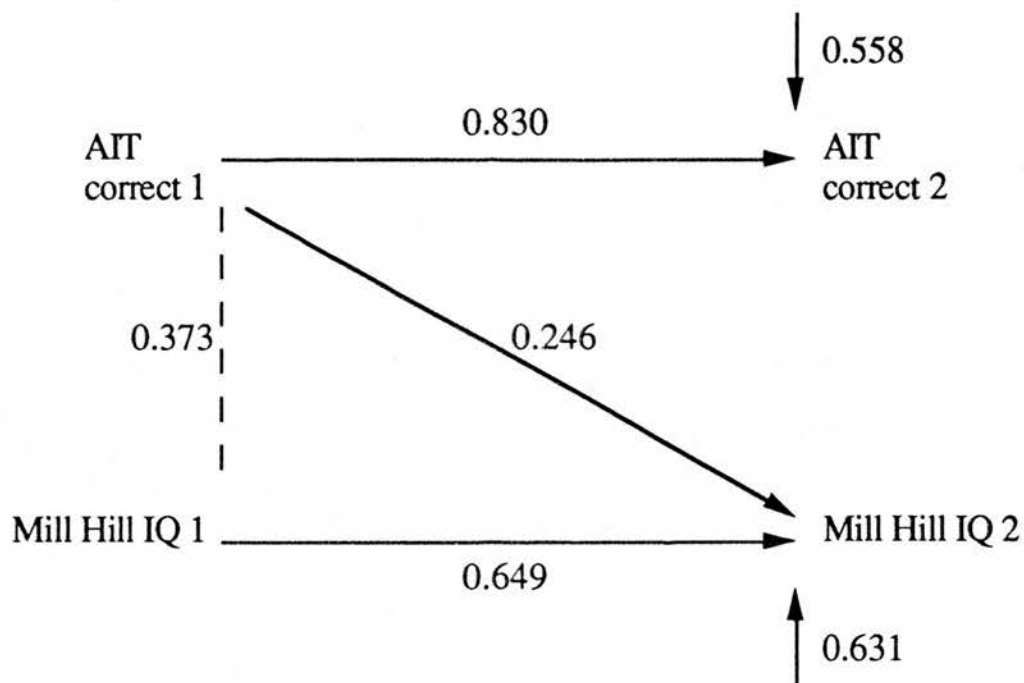
parsimony and goodness-of-fit may be applied. First, Model 1 may be preferred to Model 3. This was decided by calculating the degrees of freedom by which the models differ, i.e. 1. The difference in chi square statistics for the two models was then calculated (chi square difference between models 1 and 3 = 10.643). This chi square value at one degree of freedom may be used as a test of differential goodness-of-fit of the two models. The difference proved to be significant at  $p < 0.005$ . Model 2 ('AIT "causes" Mill Hill IQ') may be preferred to Model 1 ('Reciprocal causation') because the reduction of one degree of freedom did not improve the fit significantly (chi square difference = 0.001, d.f. = 1, ns). Moreover, the path included in Model 1, and which was omitted from Model 2, did not have a significant z statistic and, therefore, may have been fixed at zero without worsening the fit of the model. Model 3 may be considered inferior in fit to Model 2 because of the latter's poor fit indices.

Therefore Model 2 ('AIT "causes" Mill Hill IQ') was accepted. The path diagram for this model is shown in Figure 8.4a, with the parameter values adjacent to the arrows linking the variables in the model. This shows that the most acceptable interpretation of the data covariance matrix in the present study is that differences in Mill Hill IQ at time two are a result of Mill Hill IQ at time 1 and AITcorrect scores at time 1. The residual or error arrow pointing at Mill Hill IQ 2 indicates that 39.8% of the variance in measures Mill Hill IQ at time 2 was unaccounted for. The model indicates also that AITcorrect scores at time 2 were dependent to a large extent on AIT scores two years earlier but not detectably on Mill Hill IQ at time 1.

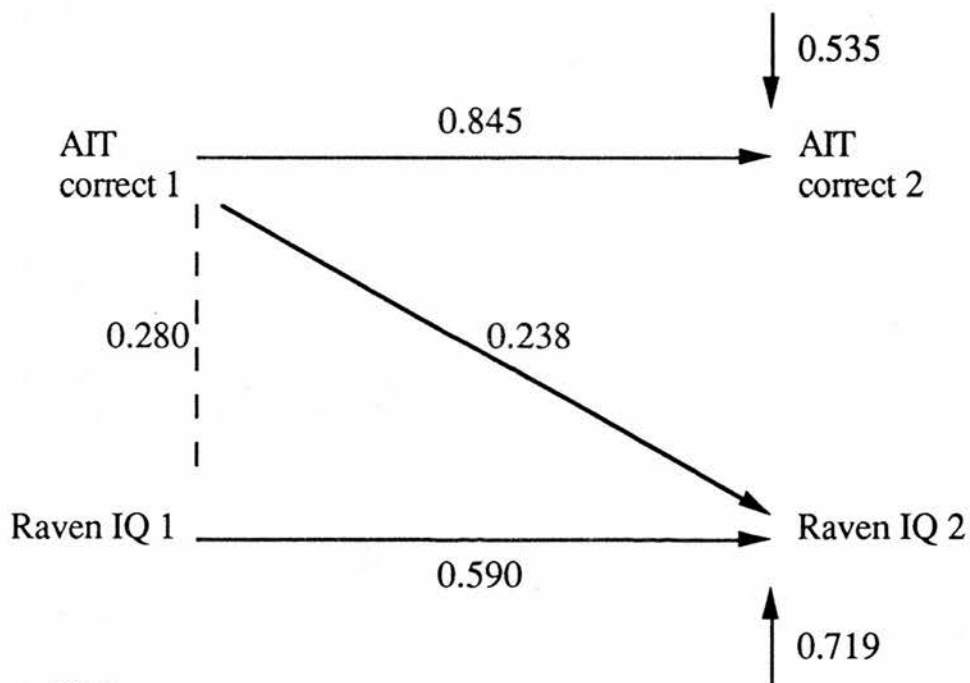
### **Models involving Raven IQ and AITcorrect scores**

Table 8.2 shows the results of the same exercise for the Raven IQ and AIT data covariance matrix in this study. The relative success of the different models was similar to that of the models tested with the Mill Hill IQ data. Model 1 ('Reciprocal causation') had high fit indices by both analysis methods, but there was a significant chi square value, indicating

a)



b)



**Figures 8.4a and 8.4b**

Optimal models from generalised least squares solutions for the relationships between: a) Mill Hill Vocabulary IQ and AITcorrect tested two years apart, and b) Raven's Progressive Matrices IQ and AIT correct tested two years apart. Dashed lines represent correlated variables. Lines with arrowheads represent causal paths in the model. Numbers adjacent to such arrows represent optimal parameter estimates in the model. Numbers adjacent to arrows without origins represent residual or error terms. Test scores at time 1 are independent variables in the model.

**Table 8.2**

Summary of structural equations program output for models of the relationships between AITcorrect and Raven IQ tested two years apart on the same subjects (n=104).

	Model 1: Reciprocal causation	Model 2: AIT 'causes' Raven IQ	Model 3: Raven IQ 'causes' AIT
<b>(i) Generalised least squares solution</b>			
Chi square (d.f.)	5.695 (1)	5.718 (2)	12.284 (2)
p value	0.017	0.057	0.002
Bentler-Bonett Normed fit index	0.998	0.998	0.995
Bentler-Bonett Non-normed fit index	0.988	0.995	0.987
Non-significant parameter values in the model	Raven 1 vs AIT 2	None	Raven 1 vs AIT 2
Z score of non- significant parameters	0.164	-	0.542
<b>(i) Maximum likelihood solution</b>			
Chi square (d.f.)	6.212 (1)	6.239 (2)	14.987 (2)
p value	0.012	0.044	<0.001
Bentler-Bonett Normed fit index	0.969	0.968	0.942
Bentler-Bonett Non-normed fit index	0.998	0.934	0.797

that the model was not acceptable by that criterion. In addition, the path linking Raven IQ at time 1 to AIT correct scores at time 2 had a non-significant z statistic, indicating that this path may have been omitted without a significant reduction in goodness-of-fit. Also, the Wald test indicated that this path might be dropped from the model without worsening its goodness-of-fit.

Model 2 ('AIT "causes" Raven IQ') had high fit indices, and chi square values from the two analysis methods that were near to conventional levels of significance. Parameters of all paths in this model had significant z scores, indicating that they all made a significant contribution to the model. Model 3 ('Raven IQ "causes" AIT') had a poor ( $<0.9$ ) Non-normed fit index on the maximum likelihood analysis, and had very highly significant chi square values by both methods of analysis. In addition, the path linking Raven IQ at time 1 to AITcorrect scores at time 2 had a non-significant z statistic. The Wald test confirmed that this path was unnecessary.

Model 2 for the Raven IQ and AITcorrect data is shown in Figure 8.4b, with path parameters adjacent to the arrows between the variables. The residual arrow pointing at Raven IQ 2 indicates that 51.6% of the variance in Raven IQ at time 2 was not accounted for after the contributions of Raven IQ at time 1 and AITcorrect scores at time 1 had been considered. Comparing Figures 8.4a and 8.4b it may be seen that AIT correct scores contribute to the variance of Raven IQ scores and Mill Hill IQ scores at time 2 to about the same extent, i.e. about 6% of the variance.

## **8.4 Discussion**

There was improvement in AIT scores over two years in the children tested here. Therefore, it may be concluded that AIT does show some maturational improvement from age 11 to age 13. However, unlike the studies reported by Nettelbeck and Wilson (1985), no control group was included to exclude the possibility that such improvements might

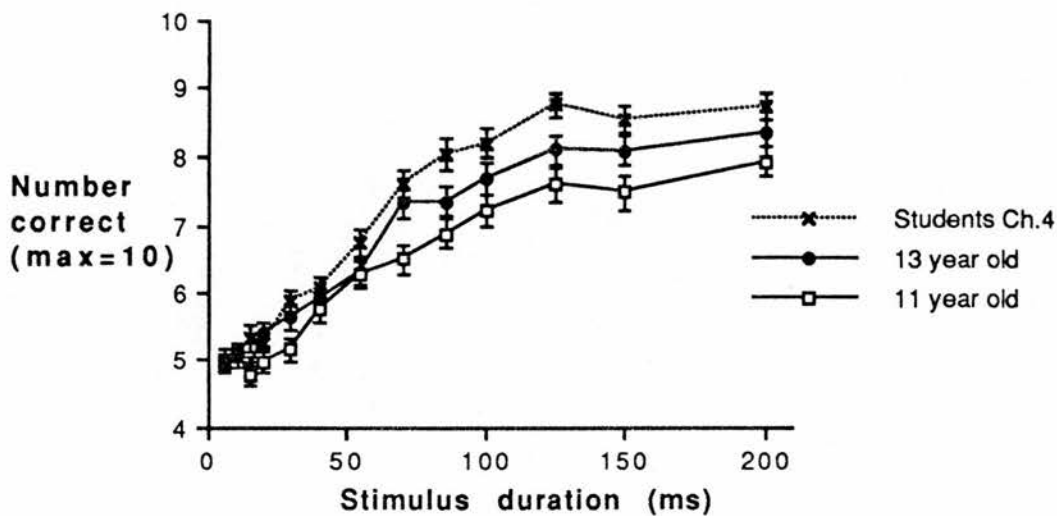
have been caused by practice on the AIT test two years previously, leading to greater familiarity on the test at the second session. Even after this possibility is allowed, it appeared that the 13-year-olds were still not performing at the level of the adult student samples. Figure 8.5 shows the schoolchildren at age 11 and 13 plotted with the 117 students who were presented in Chapter 4. The performance of the children still falls short of the students, but it should be recalled that the students are probably somewhat superior in mean IQ and, therefore, age might have been confounded with cognitive ability, in a similar fashion to the study by Nettelbeck and Wilson (1985). Of their own demonstration that the IT improvements across a period of up to two years in children, were not due to practice, Nettelbeck and Wilson (1985) opined,

This maturational change may be interpreted as arising from increasing strategic efficiency associated with general learning processes or may be attributed to increasing structural sophistication, or to some combination of both of these. ...one is limited to concluding that inspection time is subject to maturational influences and at the present time it remains an open question as to whether age differences in rate of information processing are part of the reason for this.

Therefore, even if the changes in AIT scores in the present study represent a true maturational change, they do not afford an unequivocal interpretation in terms of improvements in basic auditory processing skills. Whichever one of Nettelbeck's explanations is preferred, it seems clear that "strategic efficiency" and "structural sophistication" will be difficult to operationalise prior to testing. This limited positive evidence of a maturational change in IT with increasing age runs counter to Anderson's (1988) hypothesis that IT is a rather special information processing index which, unlike various reaction time and long term memory parameters, does not alter with age and represents a stable source of processing limitations across ages.

The present study represents one of the few longitudinal studies of IT and cognitive ability and the first involving AIT. A previous study involving visual IT in a cross-lagged panel design found no evidence to indicate that IT caused later IQ differences (Nettelbeck and





**Figure 8.5**

Mean (standard error) number of correct items on the auditory inspection time test at each of the different stimulus durations for the student sample in Chapter 4 ( $n=117$ ; cross symbol and dashed line) and the same schoolchildren tested at age 11 years in Chapter 5 ( $n=108$ ; open square symbol and continuous line) and at age 13 years for the study in this Chapter ( $n=108$ ; shaded circle symbol and continuous line). In all three samples, all subjects are included, whether or not they were classified as being able to perform the AIT task.

Young, 1990). When the AIT scores of all subjects studied over two years were included, the present study found evidence to indicate that the auditory processing abilities indexed in the AIT test caused later differences in Mill Hill and Raven IQs. The interpretation of this result might be in terms of speed of auditory processing, but AIT scores in the whole sample are likely to involve pitch discrimination skill also. The causal hypothesis was tested using a formal test of the differences between the two diagonal correlations in the cross-lagged panel and, thereafter, by constructing and testing formal structural models. The use of structural equation modelling represents an improvement over more impressionistic analyses of the cross-lagged panel. It should also be noted that, in agreement with Nettelbeck and Young (1990), when only those subjects classified as being able to perform the AIT task were included in the cross-lagged design, there was little difference in the diagonal correlations in the panel.

However, it should be noted that there are limitations in the cross-lagged panel design which is often used to offer a quasi-experimental design where formal experiments are not possible. The most influential critic of the cross-lagged design was Rogosa (1980) who stated that it was so problematic that it "is best forgotten". One main problem of the design is that differences in the reliabilities of the two variables tested at time 1 and time 2 can lead to spurious differences in the diagonal correlations. AIT reliability over two years was high in this study, by contrast with the visual IT test reliabilities of Nettelbeck and Young (1990) and those of Anderson (1986) where coefficients of 0.4 to 0.45 were typical. Another problem with the cross-lagged panel design (Humphreys and Parsons, 1979; Rogosa and Willet, 1985) is that, even after correcting for the effects of measurement error, if the differences in, say AIT changed faster between time 1 and time 2 than did the individual differences in cognitive ability test scores, then a spurious difference or equality in the cross-lagged correlations might result. These limitations should lead to caution in interpreting the present results. One solution to these limitations, suggested by Humphreys (1991), is to use up to three parallel forms of each test and to conduct the path

analyses with the latent variables underlying these test performances rather than test scores taken at a single occasion.

If the above results prove replicable, might the auditory information processing advantages be considered to have an ontogenetic relationship with cognitive ability test score differences as suggested by Brand (1984)? Perhaps faster information intake and more accurate discrimination allow more resources to be freed for the consolidation of information. In Vickers' terms (Vickers, Nettelbeck and Willson, 1972) it might be hypothesised that a child with auditory inspection time advantages need make fewer 'inspections' of a stimulus before an accurate decision is made. This would have the effect of reducing the time spent on each discrimination and might allow the person to move on to another stimulus more quickly, or allow the person more time to rehearse the already discriminated stimulus.

In summary, the present study has shown that individual differences in AIT and cognitive ability proved reliable from age 11 to age 13, and that both types of ability improved across this time. Correlations between AIT and cognitive ability were higher, around 0.4, at age 13 than at age 11. For the whole group, the auditory abilities tested in the AIT task had a significant influence on verbal and non-verbal IQ scores two years later. This represents one of the few attempts to test the hypothesis concerning the direction of causation between basic information processing skills and cognitive ability and it provided evidence to indicate that superior auditory processing skills might be causal to higher cognitive ability, as has been suggested by others (Tallal, 1989). Interestingly, a recent report by Raz, Moberg and Millman (1990) on a group of 49 subjects with a wide age range used linear modelling techniques to show that age effects on fluid intelligence were mediated by differences in frequency discrimination using 40ms tones. Therefore, basic auditory abilities which decline in age might contribute to the decline of intelligence in old age in addition to its rise in adolescence.

## Chapter Nine

### Conclusions

*The notion of elementary processing stages as the unit of intelligence is an attractive notion. It offers an alternative to the psychometric factor as the unit of analysis, a unit which seems to be afflicted with its own methodological problems. I think, though, that the initial enthusiasm for the stage unit... is now tempered as a result of a good dosage of reality testing: the anticipated correlations between stage efficiency (as measured by RT in one form or another) and IQ are very difficult to come by. ...This critique is intended as a warning to those who would embark on this course (the seasoned travellers know most of this already): it ain't easy. (Longstreth, 1984.)*

#### 9.1 Concluding remarks concerning the studies in this thesis

This Chapter has three sections. In section 9.1 the studies in this thesis will be summarised according to the themes that were investigated. In section 9.2 there will be a review of alternative auditory and visual paradigms that appear to provide approaches to sensory processing speed that future research might integrate with IT studies. Researchers in IT appear to be unaware of these paradigms and vice versa. Section 9.3 will suggest some possible reductionist, biological approaches to the further investigation of IT and the IT-cognitive ability association. Finally, section 9.4 will offer some concluding remarks.

##### 9.1.1 Characteristics of the modified AIT task

A modified AIT task was designed for the present thesis. It was designed as a modification of the original AIT task used by Brand and Deary (1982) and its modification by Nettelbeck, Edwards and Vreugdenhil (1986), in order to meet criticisms of the auditory task made by Irwin (1984). The modified AIT task resulted in threshold estimates that had durations which were higher than those of previous studies, where auditory processing speed might have been confounded with pitch discrimination ability (Chapter

4). Throughout the thesis, subjects were screened for their ability to perform the discrimination involved in the AIT task before their data were used in correlations with mental ability test scores, and a substantial minority of subjects in all studies were deemed unable to perform the task satisfactorily. This was considered to be due to their poor pitch discrimination ability skills or to the fact that some subjects had AIT thresholds at or greater than 200ms.

The AIT task appeared to produce subject performance curves that approximated to those provided by Vickers, Nettelbeck and Willson (1972) for the visual IT task (Chapter 4). When subjects' scores on the task were subjected to probit analyses, most fitted the performance model hypothesised by Vickers (1979), though competing performance models were not tested, and the individual subject data were not sufficiently detailed to allow a definite conclusion (Chapter 5). However, in possible contradiction to the accumulator model (Vickers 1979), which was formulated to account for *visual* two-choice discrimination performance, it appeared likely that chance responding on the AIT task occurred at stimulus durations significantly higher than zero milliseconds.

Various reliability estimates of the AIT task were obtained. The split-half reliability of the AIT test total score was 0.886 in a group of undergraduates (Chapter 5). The correlation between two PEST-derived AIT threshold estimates taken two weeks apart was 0.96 (Chapter 6). The correlation between two estimates of AIT threshold taken two years apart in 43 schoolchildren was 0.74, and the correlation between two measures of AIT test total score taken two years apart from 104 schoolchildren was 0.83 (Chapter 8).

Children improved in AIT performance from age 11 to age 13 years, and 13-year-olds were poorer in AIT task performance than young adults. However, the effects of practice and group differences in cognitive ability, respectively, were possible confounding effects in these results that prevented definitive conclusions being formed.

### 9.1.2 Establishing the AIT-cognitive ability association

It was argued toward the end of Chapter 1 that psychological phenomena, and the IT-cognitive ability association in particular, should be *established* first, and then attempts to *explain* the association should be mounted. The present thesis has made a series of attempts to correlate performances on the modified AIT task with mental ability test scores, which may be used to assess whether the relationship has been established (Chapters 4, 5, 6 and 7). These are summarised in Table 9.1. All correlations between AIT and estimates of verbal and non-verbal abilities in undergraduates and schoolchildren were in the expected direction, apart from a single near-zero correlation between Raven's Advanced Progressive Matrices and AIT in an undergraduate sample. The n-weighted mean correlation between AIT and verbal ability was -0.324 and between AIT and non-verbal ability was -0.269. Because of the large numbers of subjects, both of these estimates were very highly significant.

The mean raw correlations do not constitute a breakthrough the 0.3 barrier said to offer a limit to the correlations between information processing indices and measures of cognitive ability (Hunt, 1980). On the other hand, they are congruent with the sizes of the uncorrected correlations that have been found to exist between cognitive ability differences and visual IT estimates in the two largest reviews of the area to date (Nettelbeck, 1987; Kranzler and Jensen, 1987). The studies in these reviews, like the present thesis, tended to have an over-representation of undergraduates among its samples, which probably resulted in the correlations being underestimates of the 'true' values. When the student sample correlations in Table 9.1 were corrected for restriction of ability range, assuming a standard deviation of 7.5 in IQ, the mean corrected correlation between AIT and verbal tests was 0.50, and with non-verbal tests was 0.45. When correlations in the sample of 13-year-old schoolchildren were corrected for restriction of IQ range the corrected correlation between Mill Hill IQ and AIT was 0.63, and with Raven IQ was 0.47.

**Table 9.1**  
 Summary of the correlations between AIT performance indices and cognitive ability test scores in Chapters 4 to 7 of this thesis.

	n=	Subjects	Verbal test	<i>r</i> with AIT	Non-verbal test	<i>r</i> with AIT
Chapter 4	40	Undergraduates	Alice Heim 5	-0.11	Alice Heim 5	-0.40
	80	Undergraduates	Mill Hill	-0.27	Raven's advanced matrices	0.05
Chapter 5	34	Undergraduates	Alice Heim 6	-0.45	Alice Heim 6	-0.27
	53	11-year-old schoolchildren	Mill Hill	-0.36	Raven's standard matrices	-0.26
			Verbal reasoning quotient	-0.28	Mathematical reasoning quotient	-0.24
Chapter 6	21	Undergraduates	Alice Heim 6	-0.26	Alice Heim 6	-0.36
					Raven's advanced matrices	-0.14
Chapter 7	107	13-year-old schoolchildren	Mill Hill	0.42 <sup>a</sup>	Raven's standard matrices	0.41 <sup>a</sup>

<sup>a</sup>These correlations are positive because AIT test total scores were used. In all other instances, AIT thresholds (in milliseconds) were used in the correlations.



It is notable that the lowest correlations in the student samples tended to be with scores on Raven's Advanced Progressive Matrices, though this was not found for other auditory processing tasks (Raz, Willerman and Yama, 1987). However, there remains the possibility that the Raven task was particularly unsuited to uncovering an AIT-cognitive ability association. In summary, like previous reviewers of the visual IT-cognitive ability association have concluded, the association between AIT and cognitive ability is sufficiently consistent to be of theoretical interest and attempts should be made to explain it (Juhel, 1991; Levy, in press).

### **9.1.3 Explaining the AIT-cognitive ability association**

The hypothesis of Irwin (1984), that differences in AIT ability might be largely caused by individual differences in pitch discrimination ability, was tested. A novel historical review suggested that there were small but consistent correlations between sensory, particularly auditory, discrimination abilities and cognitive ability test scores or intelligence estimates, and that these were largely unknown to present writers (Chapter 3). Although significant correlations between AIT thresholds and scores on standardised pitch discrimination tasks were found, when the partial correlations between AIT and cognitive ability differences were computed, controlling for pitch discrimination ability, there were no instances where the correlations fell by a substantial amount (Chapters 5, 6 and 7). However, some evidence was found to indicate that, in schoolchildren, unspeeded pitch discrimination differences were significantly correlated with cognitive ability test scores (Chapters 5 and 7), in agreement with the historical review (Chapter 3) and more recent studies (McLeish, 1950; Lynn, Wilson and Gault, 1989).

An auditory information processing task devised by Raz, Willerman and Yama (1987), that was correlated at significant levels with cognitive ability differences in a small sample of undergraduates, appeared to offer further support for the hypothesis that pitch discrimination ability rather than auditory processing speed might form the key explanatory

variable that accounted for the AIT-cognitive ability association. However, it was argued that the so-called 'Raz' task might be testing speed of auditory processing in addition to testing pitch discrimination ability (Chapter 6). Results of two studies replicated the 'Raz' task-cognitive ability association to some degree (Chapters 6 and 7). The results of a large sample of children supported the hypothesis that 'Raz' had two sources of variance - speed of processing and discrimination - and the speed of processing aspect appeared to have a stronger association with cognitive ability test scores than the pitch discrimination aspect of the task.

A test of the hypothesis that individual differences in AIT caused IQ differences two years later, rather than the reverse, was carried out on a sample of over 100 11-year-olds retested at age 13 (Chapter 8). This followed a similar design to that of a study conducted by Nettelbeck and Young (1990) on a smaller sample of younger schoolchildren using visual IT. In order to gain a sufficient sample size to afford definite conclusions, all subjects were included, whether or not they appeared to be able to make the AIT discrimination. The results, using formal structural modelling analyses, indicated that a causal association existed between AIT and later IQ, though the problems associated with the cross-lagged panel design were discussed.

In summary, therefore, the auditory processing task used in this thesis appeared to have satisfactory psychophysical performance characteristics, individual differences on the task correlated consistently at modest levels with verbal and non-verbal ability test scores, the auditory processing ability-cognitive ability association was probably not caused by pitch discrimination differences, and AIT ability appeared to be causal to IQ differences two years later.

Following these successful attempt attempts to establish and, to some extent, explain the AIT-cognitive ability association, it is appropriate to suggest how future research in the IT

area might proceed. One obvious requirement is for a large scale study of the normal population on both auditory and visual inspection times and cognitive abilities, as suggested by Nettelbeck (1987). This may be used to establish more clearly the variance shared by IT in the two modalities and their relative strength of association with cognitive abilities. More important, perhaps, will be further attempts to 'unpack' or explain the IT-cognitive ability association. It will be suggested below that there are two types of research that will help in this regard. The first to be considered will be non-reductionistic approaches, i.e. there should be further efforts along the lines of the experiments in this thesis to include different auditory and visual tasks in the same study when examining IT-cognitive ability associations. Second, there must be reductionistic efforts, involving drug challenges and evoked potential studies, etc., to understand the biological basis of IT performance and the nature of the IT-cognitive ability association.

## **9.2 Integrating inspection time research with other paradigms**

It was suggested above that future research on IT, and on auditory and visual processing abilities in general, and cognitive ability must incorporate and consider other models of information processing performance. If this does not happen then this area of research will continue to take place with each group of researchers sticking to their own, relatively limited set of tasks. Empirical and theoretical advances may come from a more pluralistic approach.

Claims for the meanings of the correlations between experimentally-derived information processing variables and cognitive ability test scores are often based upon *a priori* assumptions about the abilities that are indexed by both types of test. It is unwise to make such claims when a single experimental test has been used in isolation. The present thesis has attempted to include more than one auditory information processing test in the same study in order to provide some convergent and discriminant validity for the hypotheses which stated that either or both of auditory processing speed and pitch discrimination

ability are correlated with cognitive ability (e.g. Chapters 6 and 7). The approach which includes two or more 'simple' processing tests in the same study may lead to a dissecting out of those basic processing abilities that correlate with mental ability type tests and those which do not. An example of such research is that of Raz, Willerman and Yama (1987), where frequency discrimination indices, but not indices from a signal detection task, were related to mental ability test scores, though both tasks appeared to offer the same degree of conscious task involvement of the subjects. The following section of this Chapter will suggest that the results of the present thesis might be extended by following this approach with other extant auditory processing tasks.

### **9.2.1 Alternative auditory information processing paradigms**

How might the results of the AIT studies reported above be integrated with other paradigms of auditory information processing? As was demonstrated in Chapter 3, there has long been an interest in asking whether sensory discrimination is related to individual differences in higher cognitive abilities. However, apart from the interest in the AIT task and the studies by Raz and colleagues (Raz, Willerman, Ingmundson and Hanlon, 1983; Raz and Willerman, 1985; Raz, Willerman and Yama, 1987), the interest in the relation between auditory capabilities and IQ-type test scores has been desultory and intermittent (Karlin, 1942; McLeish, 1950; Harris, 1964; Kalmus and Fry, 1980; Stankov and Horn, 1980). Moreover, many of the tests in these references involve auditory short-term memory tasks using spoken words, and few researchers would accept that such tasks are as 'basic' as AIT or pitch discrimination tasks. Extending the results of the AIT studies in an effort to better understand them and to integrate the meaning of the AIT task within other models of auditory processing will be difficult because of the lack of agreement about what constitute the basic, simple auditory skills.

Karlin (1942) had concluded that there were eight separate auditory abilities: auditory synthesis, auditory analysis, tonal memory, and discrimination of timbre, frequency,

intensity, duration and rhythm. Harris (1964) also reckoned that the auditory capabilities were separate, but that there were *five* of them and no general factor, and Stankov and Horn (1980) settled upon seven specific auditory processing factors and no general factor. McLeish (1950) and Elliot, Riach, Sheposh and Trahiotis (1966) went against the majority view in concluding that there existed a general auditory capability factor in addition to various specific factors, though the latter study found separate auditory discrimination and detection factors. Johnson, Watson and Jensen's review of this area of research (1987) found four relatively reliable factors from the various solutions to the numbers of separate, basic auditory abilities: auditory memory, frequency discrimination, ability to estimate loudness and duration.

However, in their own research, Johnson, Watson and Jensen (1987) began with 22 putatively basic auditory tests, though many were alternative versions of single tasks. Recently, this group (Watson, 1991) have provided a tape recorded test (the Test of Basic Auditory Capabilities; TBAC) with eight subtests: Pitch Discrimination, Single Tone Loudness, Single-Tone Duration, Pulse/Train Discrimination, Embedded Test-Tone Loudness Test, Temporal Order for Tones, Syllable Sequence Test and Syllable Identification Test. Watson (1991) reported significant correlations, ranging from 0.30 to 0.42, between the Scholastic Aptitude Test-Mathematics scores of 52 college students and scores on the TBAC, except for Pulse/Train Discrimination and Syllable Identification, where near-zero correlations were obtained. Only one of the TBAC tests correlated significantly with SAT-Verbal scores, the Embedded Test Tone Loudness Test.

The Temporal Order for Tones Test (TO) of the TBAC (Watson, 1991) is of particular interest to the studies in this thesis. It involved subjects in discriminating which of two tones, 550 Hz or 710 Hz, had been presented first. The tones were presented without a gap between them, and they were preceded and followed with a masker noise of 100ms, 625 Hz tones. The duration of the two tones was varied from 20ms to 200ms. Therefore,

all aspects of this task, except the mask noise, are similar to the AIT task, which was developed independently and from a different perspective, i.e. that of finding an analogue for the visual IT task. The correlation found by Watson (1991) between the TO task of the TBAC and SAT-Mathematics scores was 0.35, which is similar in magnitude to the uncorrected correlations reported here for the AIT task and cognitive abilities. However, the TO correlation with SAT-Verbal scores was only 0.10. A smaller study in the same report found non-significant correlations above 0.2 between SAT Mathematics and Verbal scores and the TO task in 24 learning disabled students, but higher correlations between cognitive ability differences and pitch, loudness and duration discrimination. Also, the study by Watson (1991) highlighted that *several* auditory capabilities, each thought to be independent of the other, correlated significantly with cognitive ability, indicating that the TO task and, by implication, the ability or abilities involved in the AIT task, might not have a privileged place in cognitive ability differences, though Watson's (1991) discussion does mention *discrimination* and *speed of processing* as key correlates of higher cognitive abilities. However, the statistical independence of the various auditory tests in the TBAC has not been established by using a large number of subjects (Watson, 1991; Johnson, Watson and Jensen, 1987; Watson, Johnson, Lehman, Kelly and Jensen, 1982), and Table V of the report by Johnson, Watson and Jensen (1987) indicates that a general auditory processing ability factor probably does exist. In particular, there appeared to be a close association between the TO task and frequency discrimination, which is in agreement with the results of Chapters 6 and 7 in the present thesis.

Other findings with the TO task or its forerunners appear to offer parallels with results on the AIT task (Johnson, Watson and Jensen, 1987). Some subjects on the TO task required very long durations in order to make the discrimination, despite having normal threshold values for other auditory tasks, leading the authors to conclude that,

The larger variance, relative to mean performance, for temporal-sequence-discrimination tasks, together with the large absolute values of the thresholds for some listeners on these tasks, suggest that individual differences



in this form of temporal processing may be of practical (or clinical) significance.

This conclusion is congruent with the findings of Nettelbeck, Edwards and Vreugdenhil (1986), with those of Raz and colleagues (Raz and Willerman, 1985), with those of Massaro (1976) and with the studies reported here to the effect that, for reasons that are to date obscure, some subjects require very long durations in order to make pitch discriminations for which most subjects require only short durations. TO task thresholds were very similar whether the target tones themselves were altered in duration or whether they were held at a constant duration with the gap between the tones altered.

The importance of basic auditory processing measures for school success has been suggested by Elliot, Hammer and Scholl (1989). They have developed so-called "fine-grained" auditory discrimination measures relating to speech sounds and have related these to cognitive abilities in children. They have employed synthesised consonant-vowel (ba-pa) stimuli that vary in voice onset time (VOT) from 0 to 35ms. The presence of a variable silent period (VOT) between the release of the start of the consonant sound and the beginning of voicing helps subjects to discriminate voiceless consonants, such as 'pa', from voiced consonants, such as 'ba'. The authors stated that the items of this test, "varied mainly in the temporal dimension," i.e. they indexed a form of auditory processing speed. Discrimination thresholds using an adaptive staircase technique correlated with Block Design, Coding and Digit Span (all from the WISC-R) at -0.41, -0.36 and -0.49 in 138 6- to-7 year olds. The corresponding correlations in 156 8- to 11-year-olds were -0.09, -0.21 and -0.35. The correlation between VOT thresholds and vocabulary scores was -0.57 for the younger children and -0.08 for the older children.

The results of Elliot, Hammer and Scholl (1989) appeared compatible with the results on AIT in the present study, since both focus upon speed of auditory processing. In addition, it would be interesting to investigate AIT for speech-like sounds, since these appear to be more relevant than tones to everyday auditory processing. Tallal (1989) opined that the



results of Elliot, Hammer and Scholl (1989) appeared to corroborate her own efforts over several years which had shown the importance of adequate basic auditory temporal processing speed for normal language development. However, Tallal's suggestion that the two types of auditory processing task might have arrived at similar results drew a very cold response from Elliot and colleagues (see after Tallal, 1989), and the exchange of letters degenerated into a debate about the definitions of "fine-grained", "peripheral versus central processing" and discrimination versus coding". Again, then, it should be emphasised that there is no substitute for an empirical approach; different auditory processing tasks must be included in the same studies in order to assess the degrees of variance that they share. Rational debate alone about such matters cannot lead to reliable conclusions.

Another auditory paradigm that would appear to be relevant to the speed of auditory processing indexed by the AIT task was studied by Warren and Warren (1970), Warren, Obusek, Farmer and Warren (1969) and Thomas, Hill, Carroll and Garcia (1970). This task involved the subject in discriminating the temporal order of, say, four sounds which were repeated in a loop in the same order over and over again, each element in the loop lasting for a given duration, say 200ms. Therefore, an item in such a task might have the following structure: high tone-buzz-low tone-hiss-high tone-buzz-low tone-hiss-high tone-buzz-low tone-hiss, and so on. The subject's task was to tell the tester the order of the sounds, having earlier been told which sounds will appear in the loop. When these were played at 200ms for each element undergraduates discriminated the sounds at no better than chance levels, and it required durations of between 200ms to 700ms before students were able to discriminate 50% of the loops correctly, given as long as they liked to examine the loop. Interestingly, speech sounds required a shorter stimulus duration for correct discrimination, typically between 100ms and 125ms (Thomas, Hill, Francis and Garcia, 1970).

The Warren and Warren (1970) temporal order task would appear to offer a possible

improvement over the typical IT task, in the visual and the auditory modality. This task would appear to allow a measure of speed of processing without the typical IT task's limitation of allowing the subject to hear or see the stimulus only once. In the Warren and Warren 'loop' task each stimulus element in the task masks the previous element in the loop and may be repeated for a set number of times or for as long as the subject wishes to inspect the stimulus train. Vickers has sought to construct tasks which allow subjects to sample time-limited stimuli over an extended period of time in order to accumulate information in order to make a decision, though these tasks appear to require the storage of information in short term memory and, therefore, might lead to discriminations biased by primacy and recency effects (Foreman, 1991; Vickers, Foreman, Nicholls, Innes and Gott, 1989). The Warren and Warren (1970) temporal order task has been adapted to a visual form by the present author, and the discrimination thresholds for undergraduates appear to be about 200ms, which is similar to those found in the auditory form of the task (Perry, 1991), and the thresholds derived from the task correlate above 0.5 with scores from the Wechsler Adult Intelligence Scale-Revised (Deary, unpublished results).

The original theoretical formulation of the AIT task was to the visual form of the IT task, in terms of information processing speed. However, the AIT task should also be integrated with a basic understanding of auditory abilities, as conceptualised by auditory psychophysicists. The suggestion that the AIT task might be integrated with the above auditory discrimination paradigms highlights some of the difficulties that might be met in future research along these lines. The battery of auditory tasks used by Watson (1991) takes several hours to administer, and even then the test compilers state that one is not testing adequately (Watson, Johnson and Jensen, 1987). Further, the factor structure of even this otherwise well-considered battery is unclear, and substantial covariance among tests has not been addressed. The attractive aspects of the Warren and Warren (1970) temporal order task should not obscure the fact that it presents a more complex task to the subject than the typical IT task, and it has not been integrated with other auditory

processing procedures. In summary, no agreed structure of auditory processing abilities exists which might be used to interpret straightforwardly the AIT results or, indeed, the correlations between other auditory processing indices and cognitive ability test scores. However, including further auditory tests alongside the AIT task in studies will offer some convergent and discriminant validity for hypotheses about the relation between basic information processing and higher mental abilities.

### **9.2.2 Alternative visual 'inspection time' paradigms**

It was argued above that studies attempting to explain the AIT-cognitive ability association should not be conducted in an empirical and theoretical vacuum, unaware of other approaches in auditory psychophysics, even though a monolithic understanding of auditory processing is not forthcoming from a perusal of the literature. The same may be argued for the area of visual IT. While it might be expected that other researchers in visual psychophysics would have become interested in the stimulus time required in order to make discriminations without necessarily having made specific reference to the work of Vickers, Nettelbeck and Willson (1972), it was a source of some surprise to this author to note a research effort in this field which also uses a procedure called 'inspection time', without, apparently being aware of the IT of Vickers and colleagues.

This alternative 'inspection time' procedure was devised by Bergen and Julez (1983), who investigated the ability to discriminate the presence or absence of a vertical target line embedded in an array of differently oriented lines, a task similar to that used by Triesman and Gelade (1980). Using a backward masking procedure, Bergen and Julez (1983) found that a stimulus onset asynchrony of about 60ms was required when the target line was at 90° to the other lines, but the inspection time became about 200ms when the difference was only 20°. Thus, phenomena akin to the 'inspection time' and 'noise' parameters suggested as stable individual differences in the work of Vickers, Nettelbeck and Willson (1972) were reported independently by these authors. They added, in

addition, that inspection times could be shortened by reducing the spatial area in which the target line lay. Bergen and Julez (1983) tended to support those writers in the IT field who interpreted the standard IT task as a simple rather than a higher order ability,

The process of internal scanning or searching upon which we have based much of this discussion is often described as a function of attention ...and the reader of this literature rapidly becomes aware of the fact that the word "attention" means many different things to different people. If the term is to be used to describe the phenomena described here, it should be made clear that it refers to a selective process, operating at or near the level of the primary visual cortex. ...It is, in other words, a perceptual rather than a cognitive process. The extent to which higher level control might be exercised, which would influence this process, is an open question.

Further research using the inspection time procedure of Bergen and Julez (1983) took a reductionistic approach; Zohary, Hillman and Hochstein (1990) suggested that, in parallel with the psychometric function in human visual perception whereby increases in stimulus onset asynchrony lead to more reliably correct discriminations, there exists a neural mechanism, underlying this psychological performance. This neural mechanism 'converts' the available stimulus processing time into a reduction in neuronal firing rate variance over a period of time, and variability of psychophysical performance is conceptualised by these authors in terms of the variability in firing of individual neurons. Deviations of the behavioural and physiological data appeared to be reconcilable by positing a neural network operating to make the discrimination rather than single neurons.

Another body of work which addresses processes which appear to be related to those involved in IT procedures is that of Phillips and colleagues (Phillips, 1974; Phillips and Singer, 1974; ). In this research, the stimulus onset asynchrony required in order accurately to detect stimulus onset and stimulus movement against a background of similar stimuli is estimated. Phillips has proposed neural network models for these processes and has shown that these 'simple' indices are altered in clinical states, such as chronic alcoholism (Wilson, Wiedman, Phillips and Brooks, 1988).

There is research on the early stages of visual perception which uses tachistoscopic presentation and stimulus onset asynchrony as a key variable and which has addressed: the effects of attention on line orientation and line arrangement (Cheal, Lyon and Hubbard, 1991); the relationship between the detection and identification of targets (Thomas, 1985; Broadbent and Broadbent, 1987); the effect of distractors on early perception (Kahneman, Triesman and Burkell, 1983); the effect of the locus of attention on the perception of simultaneity or otherwise of two stimuli (Stelmach and Herdman, 1991). In summary, there are several models of early visual processing which approximate more or less to the procedures and concerns of the Vickers-type visual IT paradigm. Few of these procedures have been integrated, either theoretically or empirically. As has been argued for AIT, it would appear to be appropriate to attempt to understand IT in terms of other mainstream ideas about early visual processing. The same problems of such a well-intended suggestion exist here as they did in the auditory area; the empirical effort involved in testing for individual differences in such a large number of tasks is daunting, and there is no agreed account of the psychological or physiological events involved in early visual processing.

The above discussions of alternative auditory and visual paradigms that exist for studying the early stages of perception in both modalities should not be interpreted as reducing the importance of the AIT- and visual IT-cognitive ability associations. However, they should be taken as a warning against the blithe assumption that IT-type tasks may be related to cognitive ability test scores in the belief that the former represent 'simple', 'well-understood', 'basic' psychological processes. As has been demonstrated, similar 'lower level' tasks are investigated by different research teams using different procedures and attract different, often highly technical theoretical accounts. Simple tasks are complex! One helpful way forward will be to use many more lower level processing tasks in parallel with tests of cognitive ability and use an iterative method to discover whether there are particular types of task which relate to particular cognitive abilities; such a procedure was

suggested by Jensen (1985b).

### **9.2.3 Alternative decision-making theories**

While there might be several different psychophysical procedures for investigating the early stages of visual and auditory processing, some or all of which might be relevant to pursuing further the IT-cognitive ability association, there is, in addition, no necessary agreement about the decision-making processes which underlie simple discriminations such as those involved in the IT procedure. For instance, Heath (1984) criticised the accumulator model of discrimination put forward by Vickers (1970, 1979), not least for its reliance on computer simulation methodology. Heath proposed an alternative decision-making process, the relative judgment theory, which was a modification of the random walk models of Stone (1960) and Laming (1968). Vickers' (1985) "counter-evaluation" of the relative merits of the accumulator and random walk models re-asserted his belief that the former was more adequate to explain performance on tasks such as the IT task. A full evaluation of competing decision-making theories is beyond the competence of this author; the mathematics involved in such theories is complex and often presented with large logical jumps between equations such that an attempt to follow the reasoning soon proves vain. However, the above illustration indicates that there is no agreed account of the 'basic' discrimination rules employed in even the simplest discrimination. In addition, it is not clear how the Vickers, two-counter, accumulator model or the two-threshold random walk model might be adapted to situations where one of three or more briefly presented stimuli are to be discriminated, or where the stimulus to be discriminated comes from an indeterminate number of possible target stimuli.

### **9.3 Reductionistic explanations of the IT-cognitive ability association**

An alternative to incorporating more than one information processing paradigm in order to obtain convergent and discriminant validity of the types of tasks which relate to higher cognitive abilities is to study the physiological factors which underlie the relation. Again, a



warning is required against premature reductionism. A physiological correlate of IT differences or in, say IT alterations caused by drugs, illness or aging, might have nothing at all to do with those aspects of IT performance that correlate with IQ-type test score differences. Since the IT-cognitive ability correlation at best represents about 25% of shared variance there are likely to be physiological correlates of IT performance that are task-specific and not related to IQ differences. Nevertheless, this type of research promises a more basic understanding of the AIT and visual IT abilities.

### **9.3.1 Evoked potential studies**

A promising route for obtaining a reductionistic understanding of the IT-cognitive ability association has been to obtain brain electrical evoked potential measurements concurrently as subjects perform IT task discrimination (Deary and Caryl, 1988; Caryl, 1991). Zhang, Caryl and Deary (1989a) demonstrated a significant association between IT differences and differences in the time-to-peak and the latency of the positive waveform that appeared at about 200ms (P200) after the IT stimulus onset. Over four experiments, it was demonstrated that the rise-time of the P200 wave of the EP collected during concurrent IT task performance was correlated at 0.59 ( $p < 0.001$ ) with IT test performance (Zhang, Caryl and Deary, 1989b), and it appeared that the association existed only when there was a requirement upon the subject to perform a discrimination on the stimulus. In the same study, it was found that there was a significant association between IQ and the the rise-time of the P200 wave of the EP to unmasked stimuli that required a discrimination, but not to the P200 waveform consequent upon IT stimulus onset. Zhang, Caryl and Deary (1989b) offered preliminary evidence from their EP analysis of the IT task to the effect that there might be at least three sources of variance in the visual IT task - task specific-, general ability- and confidence-related variance - and they concluded that,

the study of AEPs will contribute to the eventual understanding of the IT-IQ relationship precisely because, as we have shown above, it can provide the converging evidence that Posner (1986) argues is such a valuable adjunct to that from conventional psychological methods.



Some preliminary efforts have been made to study the evoked potential correlates of auditory inspection time test performance (Caryl, 1991). Hall's (1988) small study of undergraduates indicated that there was a significant association between temporal aspects of the P200 waveform evoked by the AIT task stimuli, but this was a small study conducted on an undergraduate sample and the results were not as clear as those reported by Zhang, Caryl and Deary (1989a,b) for the visual IT task. Evoked potential studies with the AIT task will also have to address the potential problem of stimulus presentation; whereas the two stimulus lines appear simultaneously in the visual IT task, the tones in the auditory task appear consecutively and each might evoke their own brain electrical changes. Despite the disputes which exist within the EP field concerning the interpretation of the EP waveform elements, there appears to be growing evidence that the area of the EP that occurs at 100ms to 200ms after stimulus onset is an important locus of variance in the IT task but, whether individual differences in the EP waveform during this epoch represent task-specific- or IQ-related IT variance is not known (Deary and Caryl, in press; Caryl, 1991).

### **9.3.2 Psychopharmacological and clinical studies**

A further potential source of converging evidence for understanding the IT-cognitive ability association is the study of IT and cognitive ability in certain illness states and during drug manipulations.

The study of the differences between Alzheimer's disease and Korsakoff's psychosis (Kopelman, 1985a, b) has led to a better understanding of memory mechanisms, and has demonstrated the validity of some putative components of memory to the extent that they may be shown to be damaged in one dementing disorder, but not in another. Early visual and auditory processing have been studied only a little in such groups. Schlotterer, Moscovitch and Crapper-Maclachlan (1983) showed that Alzheimer patients had a specific deficit in early visual processing when this was interrupted with backward pattern

masking, but not when contrast masking was used. Oscar-Berman, Goodglass and Cherlow (1973) had shown that Korsakoff patients were also deficient in early visual processing when this was limited by a backward pattern mask. However, both studies were limited by their poor control groups. Deary, Hunter, Langan and Goodwin (in press) have shown that Alzheimer but not Korsakoff patients have deficient IT performance when compared with sex-, age- and premorbid IQ-matched controls. Therefore, the difference in the IT performance of patients with these two brain disorders might allow some speculation about the physiological bases of IT performance, given that there are some relatively well-understood brain correlates of the cognitive decline associated with Alzheimer's and Korsakoff's disease (Deary and Whalley, 1988; Katzman and Saitoh, 1991). Hellstrom (1989) used an auditory psychophysical task involving tone duration and the 'time-order error effect' with Alzheimer patients and controls. The duration of the tones was varied and the discrimination involved tone duration, and the Alzheimer patients' deficiency on this task was interpreted in terms of limited attention capacity.

Drugs which have specific effects on brain neurotransmitter systems offer an opportunity for understanding the IT-cognitive ability association that has yet to receive much attention (Warburton, 1991). Nevertheless, studies supervised by the present author have suggested that IT is relatively robust in the face of drug insults to the brain. Petrie and Deary (1989) showed that reaction times and digit symbol task abilities were improved by smoking, but IT was unaffected. Harris (1989) showed that 1.2g oral scopolamine administered to 12 middle-aged women in a double blind crossover study caused deterioration in standard psychometric verbal and spatial ability tasks, but that verbal and visuo-spatial tachistoscopic recognition tasks using backward masking procedures were unaffected. Coull (1991) showed that decision time from the Hick reaction time task, but not movement time or inspection time were affected by acute alcohol intoxication in 16 undergraduates.

## 9.4 Conclusion

In conclusion, it must be emphasised that Longstreth's 'abandon hope all ye who enter here'-style warning that heads this Chapter is taken, but only up to a point. Chapter 1 showed that there are few well-established associations between information processing measures and cognitive ability test scores. Visual IT and auditory IT would appear to be exceptions to this. The association between cognitive ability and visual and auditory ITs is now well-established. However, its progress has not been an even one. From the excitement that IT created when it first appeared in the popular press as a putative new measure of intelligence that required little in the way of thought (Nettelbeck, 1982), it underwent a period in the doldrums when it appeared to some commentators that there might be very little association between IT and cognitive ability (Mackintosh, 1986; Howe, 1988). The late 1980s saw settling-down period, with a growing number of empirical studies and a growing consensus that the association between visual IT and cognitive ability was in the region of 0.5 (Nettelbeck, 1987; Kranzler and Jensen, 1989).

To some extent auditory IT has been a Cinderella to visual IT, having attracted much less attention. The present thesis has gone some way to redressing this balance. The results of the studies conducted here are compatible with an estimated corrected correlation of 0.5 between auditory processing speed and cognitive ability. The corrected correlation between verbal ability and AIT in children might be higher, at about 0.6. In addition, hypotheses which stated that AIT might not be indexing processing speed were tested and were found to have little to recommend them in the way of empirical results. Importantly, some steps were taken toward establishing that basic auditory processing abilities might be causal to later differences in cognitive ability.

Part of the effort of this thesis has been to show that it is helpful to be aware of precedents and parallels of IT research. The idea that basic psychological indices might account for a sizeable portion of cognitive ability variance goes back to Galton, McKeen Cattell and Spearman. We may now state that they were correct. However, it was demonstrated

above that the history of empirical tests of this idea is often retold inaccurately. Further, this thesis has demonstrated that there is a number of research efforts which appear to be indexing IT-like abilities, especially in the auditory modality, and it was argued that they should be integrated empirically and theoretically with IT in future research.

Therefore, there is much to be done to explain the robust visual and auditory IT-cognitive ability association. If the suggestions made for future research are taken up, the psychologist investigating AIT and visual IT will require skills in more than one technically demanding area of psychology. Nevertheless, they may proceed in the knowledge that one of psychology's most enduring phenomena - intelligence - is proving tractable. On the other hand, they may, with good reason, wonder whether their empirical efforts have been anticipated. One of the most surprising results obtained in the studies reported here was the association, in children at least, between cognitive ability and unspeeded pitch discrimination. Therefore, the final word will be given to Spearman (1932), and it is remarkable that, in the diamond jubilee of its publication, his *Abilities of Man* provides a strikingly up-to-date and relevant conclusion to the replication of his 1904 result in this thesis,

On the whole, the conclusion seems irresistible, that *g* is more or less involved in educing relations of likeness, even when the fundamentals are of a sensory nature.

## References

- Abelson, A.R. (1911). The measurement of mental ability of backward children. *British Journal of Psychology*, 4, 268-314.
- Ackerman, P.L. (1990). A correlational analysis of skill specificity: learning, abilities and individual differences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16, 883-901.
- Alderton, D.L., Goldman, S.R. and Pellegrino, J.W. (1985). Individual differences in process outcomes for verbal analogy and classification solution. *Intelligence*, 9, 69-85.
- Anderson, M. (1977). Mental speed and individual differences in intelligence. MA (Hons) Thesis, University of Edinburgh.
- Anderson, M. (1986). Inspection time and IQ in young children. *Personality and Individual Differences*, 7, 677-686.
- Anderson, M. (1988). Inspection time, information processing and the development of intelligence. *British Journal of Developmental Psychology*, 6, 43-57.
- Austin, J.T. and Wolfe, L.M. (1991). Annotated bibliography of structural equation modelling: technical work. *British Journal of Mathematical and Statistical Psychology*, 44, 93-152.
- Bain, A. (1868). *The Senses and the Intellect*. London: Longmans, Green and Co.
- Bagley, W.C. (1901). On the correlation of mental and motor ability in school children. *American Journal of Psychology*, 12, 193-205.
- Baltes, P.B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effect. *Human Development*, 11, 145-171.
- Barrett, P.T., Daum, I. and Eysenck, H.J. (1990). Sensory nerve conduction and intelligence: a methodological study. *Journal of Psychophysiology*, 4, 1-13.
- Barrett, P., Eysenck, H.J. and Lucking, S. (1986). Reaction time and intelligence: A replicated study. *Intelligence*, 10, 9-40.
- Beck, L.F. (1933). The role of speed in intelligence. *Psychological Bulletin*, 30, 169-178.
- Bentler, P.M. (1989). *EQS Structural Equations Program Manual*. Los Angeles, CA: BMDP Statistical Software.
- Bentler, P.M. and Speckart, G. (1981). Attitudes "cause" behaviours: a structural equation analysis. *Journal of Personality and Social Psychology*, 40, 226-238.
- Bentley, A. (1963). *Musical Ability in Children and its Measurement*. London: Harrap.
- Bergen, J.R. and Julez, B. (1983). Rapid discrimination of visual patterns. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-13, 857-863.
- Berger, M. (1982). The "scientific approach" to intelligence: an overview of its history

- with special reference to mental speed. In H.J. Eysenck (Ed.), *A Model for Intelligence*. New York: Springer.
- Bernstein, D.A., Roy, E.J., Srull, T.K. and Wickens, C.D. (1988). *Psychology*. Boston: Houghton Mifflin.
- Blaha, J. and Wallbrown, F.H. (1982). Hierarchical factor structure of the WAIS-Revised. *Journal of Consulting and Clinical Psychology*, 50, 652-660.
- Blank, G. (1934). Brauchbarkeit optischer Reaktionsmessungen. *Industrielle Psychotechnik*, 11, 140-150.
- Blinkhorn, S.F. and Hendrickson, D.E. (1982). Average evoked potentials and psychometric intelligence. *Nature*, 295, 596-597.
- Blum, L. (1978). *Pseudoscience and Mental Ability*. London: Monthly Review Press.
- Boring, E.G. (1923). Intelligence as the tests test it. *New Republic*, 35, 35-37.
- Boring, E.G. (1950). *A History of Experimental Psychology*. New York: Appleton Century Crofts.
- Bouchard, T.J., Lykken, D.T., McGue, M., Segal, N.L. and Tellegen, A. (1990). Sources of Human Psychological Differences: The Minnesota Study of Twins Reared Apart. *Science*, 250, 223-228.
- Brand, C.R. (1981). General intelligence and mental speed: their relationship and development. In J.P. Das and N. O'Connor (Eds), *Intelligence and Learning*. New York: Plenum.
- Brand, C.R. (1984). Intelligence and inspection time: an ontogenic relationship? In C.J. Turner & H.B. Miles (Eds.), *The Biology of Human Intelligence*. Hammerside: Nafferton.
- Brand, C.R. (1987a). The importance of general intelligence. In S. Modgil and C. Modgil (Eds), *Arthur Jensen: Consensus and Controversy*. Brighton: Falmer.
- Brand, C.R. (1987b). A rejoinder to Pellegrino. In S. Modgil and C. Modgil (Eds), *Arthur Jensen: Consensus and Controversy*. Brighton: Falmer.
- Brand, C.R., Caryl, P.G., Deary, I.J., Egan, V. and Pagliari, H.C. (1991). Is intelligence illusory? *Lancet*, 337, 607-608.
- Brand, C.R. and Deary, I.J. (1982). Intelligence and "inspection time". In H.J. Eysenck (Ed.), *A Model for Intelligence*. New York: Springer Verlag.
- Breitmeyer, B. (1984). *Visual Masking*. Oxford: Oxford University Press.
- Broadbent, D.E. and Broadbent, M.H.P. (1987). From detection to identification: response to multiple targets in rapid serial visual presentation. *Perception and Psychophysics*, 42, 105-113.
- Burt, C. (1909-10). Experimental tests of general intelligence. *British Journal of Psychology*, 3, 94-177.
- Bynner, J.M. and Romney, D.M. (1986). Intelligence, fact or artifact: alternative structures for cognitive abilities. *British Journal of Educational Psychology*, 56, 13-23.



- Cairns, R.B. and Ornstein, P.E. (1979). Developmental psychology. In E. Hearst (Ed.), *The First Century of Experimental Psychology*. London: LEA.
- Campbell, D.T. (1974). Evolutionary epistemology. In P.A. Schlipp (Ed.), *The Philosophy of Karl Popper, Vol. 1*. LaSalle: Open Court.
- Canavan, A.G.M., Dunn, G. and McMillan, T.M. (1986). Principal components of the WAIS-R. *British Journal of Clinical Psychology*, 25, 81-85.
- Carey, N. (1914-15). Factors in the mental processes of schoolchildren. *British Journal of Psychology*, 7, 453-490.
- Carey, N. (1915-17). Factors in the mental processes of schoolchildren. II - On the nature of specific mental factors. *British Journal of Psychology*, 8, 70-92.
- Carman, A. (1898-9). Pain and strength measurements of 1507 schoolchildren in Saginaw, Michigan. *American Journal of Psychology*, 10, 392-398.
- Carroll, J.B. (1982). The measurement of intelligence. In R.J. Sternberg (Ed.), *Handbook of Human Intelligence*. Cambridge: Viking Press.
- Caryl, P.G. (1991). Evoked potentials, inspection time and intelligence. *The Psychologist*, 4, 537-541.
- Cattell, J.McK. (1886a). The inertia of eye and brain. *Brain*, 8, 295-312.
- Cattell, J.McK. (1886b). The time taken up by cerebral operations. *Mind*, 11, 220-242, 377-392, 524-538.
- Cattell, J.McK. (1890). Mental tests and measurements. *Mind*, 15, 373-381.
- Cattell, J.McK. and Farrand, L. (1896). Physical and mental measurement of the students of Columbia University. *Psychological Review*, 3, 618-648.
- Cattell, R.B. (1963). Theory of fluid and crystallised intelligence: a critical experiment. *Journal of Educational Psychology*, 54, 1-22.
- Ceci, S.J. (1990). On the relation between microlevel processing efficiency and macrolevel measures of intelligence: some arguments against current reductionism. *Intelligence*, 14, 141-150.
- Chase, T.N., Fedio, P., Foster, N.L., Brooks, R., Dichiro, G. and Mansi, L. (1984). WAIS performance: cortical localisation by fluorodeoxyglucose F18-positron emission tomography. *Archives of Neurology*, 41, 1244-1247.
- Cheal, M., Lyon, D.R. and Hubbard, D.C. (1991). Does attention have different effects on line orientation and line arrangement discrimination? *The Quarterly Journal of Experimental Psychology*, 43A, 825-857.
- Chiang, A. and Atkinson, R.C. (1976). Individual differences and interrelationships among a select set of cognitive skills. *Memory and Cognition*, 4, 661-672.
- Child, D. (1990). *The Essentials of Factor Analysis, 2nd Edition*. London: Cassell.
- Cooper C., Kline P. and Maclaurin-Jones L. (1986). Inspection time and primary abilities. *British Journal of Educational Psychology*, 56, 304-308.



- Coull, J. (1990). The effects of alcohol and cognitive ability on human information processing and mood. MA (Hons) Thesis, University of Edinburgh.
- Crawford, J.R., Allan, K.M., Stephen, D.W., Parker, D.M. and Besson, J.A.O. (1989). The Wechsler Adult Intelligence Scale-Revised (WAIS-R): factor structure in a U.K. sample. *Personality and Individual Differences*, 10, 1209-1212.
- Cuttance, P. and Ecob, R. (Eds) (1987). *Structural Modeling by Example*. Cambridge: Cambridge University Press.
- Deary, I.J. (1980). How general is the mental speed factor in "general" intelligence? BSc (Hons) Thesis, University of Edinburgh.
- Deary, I.J. (1986). Inspection time: Discovery or rediscovery? *Personality and Individual Differences*, 7, 625-631.
- Deary, I.J. (1988a). Prospects for the biology of human intelligence. *Human Evolution*, 3, 503-513.
- Deary, I.J. (1988b). Applying evolutionary epistemology: from immunity to intelligence. *Journal of Social and Biological Structures*, 11, 399-408.
- Deary, I.J. (1988c). Positron emission tomography and the neurobiology of intelligence. *Mensa Research Journal*, 24, 10-12.
- Deary, I.J. (In press). Inspection time and WAIS-R intelligence subtypes: a confirmatory factor analysis study. *Intelligence*.
- Deary, I.J. and Caryl, P.G. (1988). Inspection time and cognitive ability. In K.J. Gilhooly, M.T.G. Keane, R.H. Logie and G. Erdos (Eds), *Lines of Thinking: Reflections on the Psychology of Thought. Volume 2: Skills, Emotion, Creative Processes, Individual Differences and Teaching Thinking*. Chichester: Wiley.
- Deary, I.J. and Caryl, P.G. (In press). Intelligence, EEG and evoked potentials. In P.A. Vernon (Ed.), *Biological Approaches to Intelligence*. Norwood, N.J.: Ablex.
- Deary, I.J., Hunter, R., Langan, S.J. and Goodwin, G.M. (In press). Inspection time, psychometric intelligence and clinical estimates of cognitive ability in presenile Alzheimer's disease and Korsakoff's psychosis. *Brain*.
- Deary, I.J., Langan, S.J., Graham, K.S., Hepburn, D.A. and Frier, B.M. (In press). Severe recurrent hypoglycaemia, intelligence and speed of information processing. *Intelligence*.
- Deary, I.J. and Whalley, L.J. (1988). Recent research on the causes of Alzheimer's disease. *British Medical Journal*, 297, 807-810.
- Eckberg, D.L. (1979). *Intelligence and Race*. New York: Praeger.
- Ecob, R. (1987). Applications of structural equation modeling to longitudinal educational data. In P. Cuttance and R. Ecob (Eds), *Structural Modeling by Example*. Cambridge: Cambridge University Press.
- Edelman, G.M. and Reeke, G.N. (1982). Selective networks capable of representative transformations, limited generalisations and associative memory. *Proceedings of the National Academy of Sciences of the USA*, 79, 2091-2095.

- Edwards C. (1984). Inspection time in three sensory modalities and its relation to intelligence. B.A. (Hons) Thesis, University of Adelaide.
- Egan V. (1986). Intelligence and inspection time: Do high IQ subjects use cognitive strategies? *Personality and Individual Differences*, 7, 695-700.
- Egan, V. (1991). Intelligence, inspection time and cognitive strategies. PhD Thesis, University of Edinburgh.
- Egan, V. and Deary, I.J. (In press). Are specific inspection time strategies prevented by concurrent tasks? *Intelligence*.
- Elliot, L.L., Hammer, M.A. and Scholl, A.E. (1989). Fine-grained auditory discrimination in normal children and children with language-learning problems. *Journal of Speech and Hearing Research*, 32, 112-119.
- Elliot, D.N., Riach, W.D., Sheposh, J.P. and Trahiotis, C. (1966). Discrimination of high school sophomores on a battery of auditory tests. *Acta Otolaryngologica Supplement*, 216, 1-59.
- Ertl, J. and Schafer, E.W.P. (1969). Brain response correlates of psychometric intelligence. *Nature*, 223, 421-422.
- Eysenck, H.J. (1967). Intelligence assessment: a theoretical and experimental approach. *British Journal of Educational Psychology*, 37, 81-98.
- Eysenck, H.J. (1973). *The Inequality of man*. London: Temple Smith.
- Eysenck, H.J. (1979). *The Structure and Measurement of Intelligence*. New York: Springer.
- Eysenck, H.J. (Ed.) (1982). *A Model for Intelligence*. New York: Springer.
- Eysenck, H.J. (1986). Toward a new model of intelligence. *Personality and Individual Differences*, 7, 731-736.
- Fancher, R.E. (1985a). Spearman's original computation of g: A model for Burt? *British Journal of Psychology*, 76, 341-352.
- Fancher, R.E. (1985b) *The Intelligence Men*. London: Norton.
- Finney, D.J. (1971). *Probit Analysis*. Cambridge: Cambridge University Press.
- Forrest, D.W. (1974). *Francis Galton: The Life and Work of a Victorian Genius*. London: Elek.
- Foreman, E. (1991). Individual differences in expanded judgment tasks. PhD Thesis, University of Adelaide.
- Frearson, W., Barrett, P.T. and Eysenck, H.J. (1988). Intelligence, reaction time and the effects of smoking. *Personality and Individual Differences*, 9, 497-518.
- Galton, F. (1883). *Inquiries into Human Faculty*. London: Dent.
- Giannitrapani, D. (1971). Scanning mechanisms and the EEG. *Electroencephalography and Clinical Neurophysiology*, 30, 139-146.

- Gilbert, D.G., Johnson, S., Gilbert, B.O. and McColloch, M.A. (1991). Event-related potential correlates of IQ. *Personality and Individual Differences*, 12, 1183-1184.
- Godfrey Thomson Unit. (1969). *Mathematics 4A*. London: University of London Press.
- Godfrey Thomson Unit. (1970). *Verbal Reasoning Test 89*. London: University of London Press.
- Gould, S.J. (1981). *The Mismeasure of Man*. New York: Norton.
- Grieve, R. (1979). Inspection time and intelligence. MA (Hons) Thesis, University of Edinburgh.
- Griffing, H. (1895-6). On the development of visual perception and attention. *American Journal of Psychology*, 7, 227-238.
- Guilford, J.P. (1985). The structure of intellect model. In B.B. Wolman (Ed.), *Handbook of Intelligence*. Chichester: Wiley.
- Gustafsson, J.E. (1984). A unifying model for the structure of mental abilities. *Intelligence*, 8, 179-203.
- Haier, R.J., Robinson, D.L., Braden, W. and Williams, D. (1983). Electrical potentials of the cortex and psychometric intelligence. *Personality and Individual Differences*, 4, 591-599.
- Haier, R.J., Siegel, B.V., Neuchterlein, K.H., Hazlett, E., Wu, J.C., Paek, J., Browning, H.L. and Buchsbaum, M. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence*, 12, 199-217.
- Hakstian, A.R. and Cattell, R.B. (1982). *Manual for the Comprehensive Ability Battery (CAB)*. Champaign, IL: IPAT.
- Hall, B. (1988). Evoked potentials, auditory inspection time and intelligence. M.A.(Hons) Thesis, University of Edinburgh.
- Harris, J.D. (1964). A factor analytic study of three signal detection abilities. *Journal of Speech and Hearing Research*, 7, 71-78.
- Harris, M.B. (1989). Effects of oral scopolamine on lateralized cerebral function. MA (Hons) Thesis, University of Edinburgh.
- Hayek, F.A. (1952). *The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology*. London: Routledge, Kegan and Paul.
- Hearnshaw, L. (1979). *Cyril Burt: Psychologist*. New York: Cornell University Press.
- Heath, R.A. (1984). Random walk and accumulator models of psychophysical discrimination: a critical evaluation. *Perception*, 13, 57-65.
- Hebb, D.O. (1949). *The Organization of Behaviour*. London: Chapman and Hall.
- Heim A. (1968). Manual for the AH5 Group Test of High Grade Intelligence. NFER-Nelson, Windsor (Revised Edition).

Heim, A.W., Watts, K.P. & Simmonds, V. (1983). *AH6 Group Tests of High Level Intelligence Manual*. Windsor: NFER-Nelson.

Hellstrom, A. (1989). Discrimination optimization: a governing principle for psychophysics. In D. Vickers and P.L. Smith (Eds), *Human Information Processing: Measures, Mechanisms and Models*. Amsterdam: North Holland.

Hendrickson, D.E. and Hendrickson, A.E. (1980). The biological basis of individual differences in intelligence. *Personality and Individual Differences*, 1, 1-33.

Hendrickson, A.E. and Hendrickson, D.E. (1982). The biological basis of intelligence: Parts I and II. In H.J. Eysenck (Ed.), *A Model for Intelligence*. New York: Springer Verlag.

Herrnstein, R.J. and Boring, E.G. (1965). *A Source Book in the History of Psychology*. Cambridge MA: Harvard University Press.

Hick, W.E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4, 11-26.

Hooper, F.H., Hooper, J.O. and Colbert, K.K. (1984). *Personality and memory correlates of intellectual functioning: young adulthood to old age*. London: Karger.

Horn, J.L. (1980). Concepts of intellect in relation to learning and adult development. *Intelligence*, 4, 285-317.

Howe, M.J.A. (1988a). Intelligence as an explanation. *British Journal of Psychology*, 79, 349-360.

Howe, M.J.A. (1988b). The hazards of using correlational evidence as a means of identifying the causes of individual ability differences: a rejoinder to Sternberg and a reply to Miles. *British Journal of Psychology*, 79, 539-545.

Howe, M.J.A. (1990). *The Origins of Exceptional Abilities*. Oxford: Blackwell.

Humphreys, L.G. (1991). Causal inferences from observational data: use a redesigned cross-lagged methodology. *Intelligence*, 15, 151-156.

Humphreys, L.G. and Parsons, C.K. (1979). A simplex process model for describing differences between cross-lagged correlations. *Psychological Bulletin*, 86, 325-334.

Hunt, E. (1980). Intelligence as an information processing concept. *British Journal of Psychology*, 71, 449-474.

Hunt, E.B., Lunneborg, C. & Lewis, J. (1975). What does it mean to be high verbal? *Cognitive Psychology*, 7, 194-227.

Hunter, J.E. and Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.

Inouye, E, Park, K.S. and Asaka, A. (1984). Blood uric acid level and IQ: a study in twin families. *Genet. Med. Gemellol.*, 33, 237-242.

Irwin, R.J. (1984). Inspection time and its relation to intelligence. *Intelligence*, 8, 47-65.

Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39, 1-123.

- Jensen, A.R. (1980). *Bias in Mental Testing*. London: Methuen.
- Jensen, A.R. (1982). Reaction time and psychometric g. In H.J. Eysenck (Ed.), *A Model for Intelligence*. New York: Springer.
- Jensen, A.R. (1984). Test validity: g versus the specificity hypothesis. *Journal of Social and Biological Structures*, 7, 93-118.
- Jensen, A.R. (1985a). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *Behavioural and Brain Sciences*, 8, 193-263.
- Jensen, A.R. (1985b). The plasticity of intelligence at different levels of analysis. In J. Lockhead et al. (Eds), *Thinking: Progress in Research and Teaching*. Philadelphia: Franklin Institute Press.
- Jensen, A.R. (1987a). Process differences and individual differences in some cognitive tasks. *Intelligence*, 11, 107-136.
- Jensen, A.R. (1987b). Individual differences in the Hick paradigm. In P.A. Vernon (Ed.), *Speed of Information-Processing and Intelligence*. Norwood, N.J.: Ablex.
- Jensen, A.R. and Munro, E. (1979). Reaction time, movement time and intelligence. *Intelligence*, 3, 121-126.
- Jensen, A.R. & Vernon, P.A. (1986). Jensen's reaction-time studies: a reply to Longstreth. *Intelligence*, 10, 153-179.
- Johnson, D.M., Watson, C.S. and Jensen, J.K. (1987). Individual differences in auditory capabilities. I. *Journal of the Acoustical Society of America*, 81, 427-438.
- Joreskog, K.G. and Sorbom, D. (1986). *LISREL: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. User's Guide*. Mooresville, Indiana: Scientific Software, Inc.
- Joynson, R.B. (1989). *The Burt Affair*. London: Routledge.
- Juhel, J. (1991). Relationships between psychometric intelligence and information-processing speed indexes. *Cahiers de Psychologie Cognitive / European Bulletin of Cognitive Psychology*, 11, 73-105.
- Kahneman, D., Treisman, A. and Burkell, J. (1983). The cost of visual filtering. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 510-522.
- Kaiser, H.F. (1962). Image analysis. In C.W. Harris (Ed.), *Problems in Measuring Change*. Wisconsin: University of Wisconsin Press.
- Kallman, H.J., Hirtle, S.C. and Davidson, D. (1986). Recognition masking of auditory duration. *Perception and Psychophysics*, 40, 45-52.
- Kallman, H.J. and Massaro, D.W. (1979). Similarity effect in backward recognition masking. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 110-128.
- Kalmus, H. and Fry, D.B. (1980). On tune deafness (dysmelodia): frequency, development, genetics and musical background. *Annals of Human Genetics*, 43, 369-382.



- Kamin, L.J. (1974). *The Science and Politics of IQ*. London: LEA.
- Karlin, J.E. (1942). A factorial study of auditory function. *Psychometrika*, 7, 251-279.
- Katzman, R. and Saitoh, T. (1991). Advances in Alzheimer's disease. *FASEB Journal*, 5, 278-286.
- Kline, P. (1991a). *Intelligence: The Psychometric View*. London: Routledge.
- Kline, P. (1991b). Sternberg's components: non-contingent concepts. *Personality and Individual Differences*, 12, 873-876.
- Kopelman, M.D. (1985a). Rates of forgetting in Alzheimer-type dementia and Korsakoff's syndrome. *Neuropsychologia*, 23, 623-638.
- Kopelman, M.D. (1985b). Multiple memory deficits in Alzheimer-type dementia: implications for pharmacotherapy. *Psychological Medicine*, 15, 527-541.
- Kranzler, J.H. and Jensen, A.R. (1989). Inspection time and intelligence: a meta-analysis. *Intelligence*, 13, 329-347.
- Lally, M. and Nettelbeck, T. (1977). Intelligence, reaction time and inspection time. *American Journal of Mental Deficiency*, 82, 273-281.
- Laming, D.R.J. (1968). *Information Theory of Choice-Reaction Times*. London: Academic Press.
- Lancet (1991). The illusion of intelligence. *Lancet*, 337, 208.
- Lee, S.-Y. (1985). On testing functional constraints in structural equation models. *Biometrika*, 72, 125-131.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustic Society of America*, 49, 467-477.
- Levy, P. (In press). Inspection time and its relation to intelligence: issues of measurement and meaning. *Personality and Individual Differences*.
- Liberson, W.T. (1989). Contribution to the hypothesis that one of the functions of brain waves is the frequency analysis of incoming stimuli. *Journal of Clinical Neurophysiology*, 6, 292 (Abstract).
- Link, S.W. and Heath, R.A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, 40, 77-105.
- Livson, N. and Krech, D. (1956). Dynamic systems, perceptual differentiation and intelligence. *Journal of Personality*, 25, 46-58.
- Longstreth, L.E. (1984). Jensen's reaction time investigations of intelligence: a critique. *Intelligence*, 8, 139-160.
- Lynn, R., Wilson, R.G. and Gault, A. (1989). Simple musical tests as measures of Spearman's g. *Personality and Individual Differences*, 10, 25-28.
- Longstreth, L.E., Walsh, D.A., Alcorn, M.B., Szeszulski, P.A. and Manis, F.R. (1986). Backward masking, IQ, S.A.T. and reaction time: Interrelationships and theory.

*Personality and Individual Differences*, 7, 643-651.

Lubin, M.-P. and Fernandez, J.M. (1986). The relationship between psychometric intelligence and inspection time. *Personality and Individual Differences*, 7, 653-657.

Mackenzie, B. and Bingham, E. (1985). IQ, inspection time and response strategies in a university population. *Australian Journal of Psychology*, 37, 257-268.

Mackenzie, B. & Cumming, S. (1986). How fragile is the relationship between inspection time and intelligence: the effects of apparent motion cues and previous experience. *Personality and Individual Differences*, 7, 721-729.

Mackenzie, B., Molloy, E., Martin, F., Lovegrove, W. and McNicol, D. (1991). Inspection time and the content of simple tasks: a framework for research on speed of information processing. *Australian Journal of Psychology*, 43, 37-43.

Mackintosh, N.J. (1981). A new measure of intelligence? *Nature*, 289, 529-530.

Mackintosh, N.J. (1986). The biology of intelligence? *British Journal of Psychology*, 77, 1-18.

McDougall, W. (1905). *Physiological Psychology*. London: Dent.

McLeish, J. (1950). The validation of Seashore's measures of musical talent by factorial method. *British Journal of Psychology-Statistical Section*, 3, 129-140.

McNemar Q. (1955). *Psychological Statistics (2nd Edn)*. Wiley, London.

Massaro, D.W. (1970). Perceptual auditory images. *Journal of Experimental Psychology*, 85, 411-417.

Massaro, D.W. (1972). Perceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 79, 124-145.

Massaro, D.W. (1973). A comparison of forward versus backward recognition masking. *Journal of Experimental Psychology*, 100, 434-436.

Massaro, D.W. (1976). Auditory information processing. In W.K. Estes (Ed.), *Handbook of Learning and Cognitive Processes Vol. 4*. Hillsdale, NJ: LEA.

Massaro D.W. and Kahn, B.J. (1973). Effects of central processing on auditory recognition. *Journal of Experimental Psychology*, 97, 51-58.

Miles, T.R. (1988). Comments on Howe's paper. *British Journal of Psychology*, 79, 535-538.

Miles, W.R. and Bell, H.M. (1929). Eye movement records in the investigation of study habits. *Journal of Experimental Psychology*, 12, 450-458.

Mulholland, T.M., Pellegrino, J.W. and Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology*, 12, 252-284.

Muise, J.G., LeBlanc, R.S., Lavoie, M.E. and Arsenault, A.S. (1991). Two-stage model of visual backward masking: sensory transmission and accrual of effective information as a function of target intensity and similarity. *Perception and Psychophysics*, 50, 197-204.



- Nettelbeck, T. (1982). Inspection time: an index for intelligence? *Quarterly Journal of Experimental Psychology*, 34A, 299-312.
- Nettelbeck, T. (1987). Inspection time and intelligence. In P.A. Vernon (Ed.), *Speed of Information Processing and Intelligence*. Norwood N.J.: Ablex.
- Nettelbeck T., Edwards C. and Vreugdenhil A. (1986). Inspection time and IQ: Evidence for a mental speed-ability association. *Personality and Individual Differences*, 7, 633-641.
- Nettelbeck T. and Lally M. (1976). Inspection time and measured intelligence. *British Journal of Psychology*, 67, 17-22.
- Nettelbeck, T. and Wilson, C. (1985). A cross-sequential analysis of developmental differences in speed of visual information processing. *Journal of Experimental Child Psychology*, 40, 1-22.
- Nettelbeck, T. and Young, R. (1990). Inspection time and intelligence in 7-yr-old children: a follow-up. *Personality and Individual Differences*, 11, 1283-1290.
- O'Grady, K.E. (1989). Factor structure of the WISC-R. *Multivariate Behavioural Research*, 24, 177-193.
- Oscar-Berman, M., Goodglass, H. and Cherlow, D.G. (1973). Perceptual laterality and iconic recognition of visual materials by Korsakoff patients and normal adults. *Journal of Comparative and Physiological Psychology*, 82, 316-321.
- Peak H. and Boring, E.G. (1926). The factor of speed in intelligence. *Journal of Experimental Psychology*, 9, 71-94.
- Perry, A. (1991). Detection of temporal order in briefly presented letter sequences: effects of item duration and repeated presentation and correlation with inspection time. MA (Hons) Thesis, University of Edinburgh.
- Petrie, R.X.A. and Deary, I.J. (1989). Smoking and human information processing. *Psychopharmacology*, 99, 393-396.
- Phillips, W.A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics*, 16, 283-290.
- Phillips, W.A. and Singer, W. (1974). Function and interaction of on- and off-transients in vision. I. Psychophysics. *Experimental Brain Research*, 19, 493-506.
- Piaget, J. (1971). *Biology and Knowledge*. Edinburgh: Edinburgh University Press.
- Piaget, J. (1978). *Behaviour and Evolution*. London: Routledge, Kegan and Paul.
- Piaget, J. (1980). *Adaptation and Intelligence*. Chicago: University of Chicago Press.
- Plomin, R. and Rende, R. (1991). Human behavioural genetics. *Annual Review of Psychology*, 42, 161-190.
- Plotkin, H.C. (1987). Evolutionary epistemology as science. *Biology and Philosophy*, 2, 295-313.
- Puckett, J.M. and Kausler, D.H. (1984). Individual differences and models of memory span: a role for memory search rate? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 72-82.

- Raven, J.C. (1958). *Standard Progressive Matrices*. London: Lewis.
- Raven J.C., Court J.H. and Raven J. (1977). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Lewis, London.
- Raven, J.C., Raven, J. & Court, J.H. (1982). *Mill Hill Vocabulary Scale*. London: Lewis.
- Raz, N., Moberg, P.J. and Millman, D. (1990). Effects of age and age-related differences in auditory information processing on fluid and crystallised intelligence. *Personality and Individual Differences*, 11, 1147-1152.
- Raz, N. and Willerman, L. (1985). Aptitude-related differences in auditory information processing: effects of selective attention and tone duration. *Personality and Individual Differences*, 6, 299-304.
- Raz, N., Willerman, L., Ingmundson, P. and Hanlon, M. (1983). Aptitude-related differences in auditory recognition masking. *Intelligence*, 7, 71-90.
- Raz, N., Willerman, L. & Yama, M. (1987). On sense and senses: intelligence and auditory information processing. *Personality and Individual Differences*, 8, 201-210.
- Rhodes, L., Dustman, R. and Beck, E. (1969). The visual evoked response: a comparison of bright and dull children. *Electroencephalography and Clinical Neurophysiology*, 27, 364-372.
- Rogosa, D. (1980). A critique of cross-lagged correlations. *Psychological Bulletin*, 88, 245-258.
- Rogosa, D. and Willett, J.B. (1985). Satisfying a simplex structure is simpler than it should be. *Journal of Educational Statistics*, 10, 99-107.
- Rose, S., Lewontin, R.J. and Kamin, L.J. (1984). *Not in Our Genes*. Pelican Books.
- Ross, L.E. and Ward, T.B. (1978). The processing of information from short-term visual store: developmental and intellectual level differences. In N.R. Ellis (ED.), *International Review of Research into Mental Retardation Vol 9*. New York: Academic Press.
- Roth, E. (1964). Die Geschwindigkeit der Verarbeitung von Information und ihr Zusammenhang mit Intelligenz. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 11, 616-622.
- Rushton, J.P. (1991). Mogoloid-caucasoid differences in brain size from military samples. *Intelligence*, 15, 351-359.
- Saccuzzo, D.P., Larson, G.E. and Rimland, B. (1986). Visual, auditory and reaction time approaches to the measurement of speed of information processing and individual differences in intelligence. *Personality and Individual Differences*, 7, 659-667.
- Satorra, A. (1989). Alternative test criteria in covariance structure analysis: a unified approach. *Psychometrika*, 54, 131-151.
- Schlotterer, G., Moscovitch, M. and Crapper-McLachlan, D. (1983). Visual processing deficits as assessed by spatial frequency contrast sensitivity and backward masking in normal ageing and Alzheimer's disease. *Brain*, 107, 309-325.

- Schucard, D.W. and Horn, J.L. (1972). Evoked cortical potentials and measurement of human abilities. *Journal of Comparative and Physiological Psychology*, 78, 59-68.
- Schwartzman, A.E., Gold, D., Andres, D., Arbuckle, T.Y. and Chaikelson, J. (1987). Stability of intelligence: a 40-year follow-up. *Canadian Journal of Psychology*, 41, 244-256.
- Seashore, C.E., Lewis, D. & Saetveit, J.G. (1956). *Seashore Measures of Musical Talents*. New York: Psychological Corporation.
- Shallice, T. (1964). The detection of change and the perceptual moment hypothesis. *British Journal of Statistical Psychology*, 17, 113-135.
- Sharp, S.E. (1898-9). Individual psychology: A study in psychological method. *American Journal of Psychology*, 10, 329-391.
- Sinet, P.M., Lejeune, J. and Jerome, H. (1979). Trisomy 21 (Down's syndrome), glutathione peroxidase, hexose monophosphate shunt and IQ. *Life Sciences*, 24, 29-33.
- Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Spearman, C. (1932). *The Abilities of Man*. London: Macmillan.
- SPSS (1990). *SPSS Advanced Statistics User's Guide*. Chicago, IL: SPSS Inc.
- Stankov, L. and Horn, J.L. (1980). Human abilities revealed through auditory tests, *Journal of Educational Psychology*, 72, 21-44.
- Steiger, J.H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245-251.
- Stelmach, L.B. and Herdman, C.M. (1991). Directed attention and the perception of temporal order. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 539-550.
- Sternberg, R.J. (1977). Component processes in analogical reasoning. *Psychological Review*, 84, 353-378.
- Sternberg, R.J. (1981). Nothing fails like success: the search for an intelligent paradigm for studying intelligence. *Journal of Educational Psychology*, 73, 142-155.
- Sternberg, R.J. (1985a). Human intelligence: the model is the message. *Science*, 230, 1111-1118.
- Sternberg, R.J. (1985b). *Beyond IQ: a triarchic theory of human intelligence*. Cambridge: Cambridge University Press.
- Sternberg, R.J. (1988). Explaining away intelligence: a reply to Howe. *British Journal of Psychology*, 79, 527-534.
- Sternberg, R.J. (1990). *Metaphors of Mind: Conceptions of the Nature of Intelligence*. Cambridge: Cambridge University Press.
- Sternberg, R.J. and Gardner, M.K. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General*, 112, 80-116.

- Sternberg, S. (1966). High speed scanning in human memory. *Science*, 153, 652-654.
- Sternberg, S. (1969). Memory-scanning: mental processes revealed by reaction-time experiments. *American Scientist*, 57, 421-457.
- Stillman, J.A. (1989). A comparison of three adaptive psychophysical procedures using inexperienced listeners. *Perception and Psychophysics*, 46, 345-350.
- Stone, M. (1960). Models for reaction time. *Psychometrika*, 25, 251-260.
- Stough, C.K.K., Nettelbeck, T. and Cooper, C.J. (1990). Evoked brain potentials, string length and intelligence. *Personality and Individual Differences*, 11, 401-406.
- Sugarman, L. (1986). *Life-Span Development: Concepts, Theories and Interventions*. London: Methuen.
- Tallal, P. and Piercy, M. (1973). The repetition test. *Neuropsychologia*, 11, 389-398.
- Tallal, P. (1989). Fine-grained discrimination deficits in language-learning impaired children are specific neither to the auditory modality nor to speech perception. *Journal of Speech and Hearing Research*, 32, 616-619.
- Tallal, P., Stark, R.E. and Mellits, E.D. (1985a). The relationship between auditory temporal analysis and receptive language development: evidence from studies of developmental language disorder. *Neuropsychologia*, 23, 527-534.
- Tallal, P., Stark, R.E. and Mellits, E.D. (1985b). Identification of language-impaired children on the basis of rapid perception and production skills. *Brain and Language*, 25, 314-322.
- Taylor M.M. and Creelman C.D. (1967). PEST: Efficient estimates in probability functions. *Journal of the Acoustical Society of America*, 4, 782-787.
- Thomas, I.B., Hill, P.B., Carroll, F.S. and Garcia, B. (1970). Temporal order in the perception of vowels. *Journal of the Acoustical Society of America*, 48, 1010-1013.
- Thomas, J.P. (1985). Detection and identification: how are they related? *Journal of the Optical Society of America*, 2, 1457-1467.
- Thomson, G.H. (1939). *The Factorial Analysis of Human Ability*. London: Houghton Mifflin.
- Thorndike, E.L., Lay, W. and Dean, P.R. (1909). The relation of accuracy in sensory discrimination to general intelligence. *American Journal of Psychology*, 20, 364-369.
- Thurstone, L.L. (1938). Primary mental abilities. *Psychometric Monographs No. 1*. Chicago: University of Chicago Press.
- Todman, J. and Gibb, C.McD. (1985). High speed memory scanning in retarded and non-retarded adolescents. *British Journal of Psychology*, 76, 49-57.
- Turvey, M.T. (1973). On peripheral and central processes in vision: inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, 80, 1-52.
- Travis, E.L. and Hunter, T.A. (1928). The relation between "intelligence" and reflex



conduction rate. *Journal of Experimental Psychology*, 11, 342-354.

Treisman, A.M. and Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.

Undheim, J.O. (1981a). On intelligence II: a neo-Spearman model to replace Cattell's theory of fluid and crystallised intelligence. *Scandinavian Journal of Psychology*, 22, 181-187.

Undheim, J.O. (1981b). On intelligence IV: toward a restoration of general intelligence. *Scandinavian Journal of Psychology*, 22, 251-265.

Undheim, J.O. and Gustafsson, J.E. (1987). The hierarchical organisation of cognitive abilities: restoring general intelligence through the use of linear structural relations. *Multivariate Behavioural Research*, 22, 149-171.

Vernon, P.A. (1986). Inspection time: Does it measure intelligence. *Personality and Individual Differences*, 7, 715-720.

Vernon, P.A. (Ed.) (1987). *Speed of information processing and intelligence*. Norwood, N.J.: Ablex.

Vernon, P.A., Weese, S.A. and Miller, L.T. (1991). Is Spearman's "g" a unitary construct? Evidence suggesting not. Paper presented to the International Society for the Study of Individual Differences 5th Biennial Conference, Oxford, U.K., July.

Vernon, P.E. (1940). *The Measurement of Abilities*. London: University of London Press.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Personality and Individual Differences*, 7, 609-623.

Vickers, D. (1979). *Decision Processes in Visual Perception*. London: Academic Press.

Vickers, D., Foreman, E.A., Nicholls, M.E.A., Innes, N.J. and Gott, R.E. (1989). Some experimental tests of the application of an interval of uncertainty model to a time-limited expanded judgment task. In D. Vickers and P.L. Smith (Eds), *Human Information Processing: Measures, Mechanisms and Models*. Amsterdam: North Holland.

Vickers, T., Nettelbeck, T. & Willson, R.J. (1972). Perceptual indices of performance: the measurement of 'inspection time' and 'noise' in the visual system. *Perception*, 1, 263-295.

Vickers, D. and Smith, P.L. (1985). Accumulator and random walk models of psychophysical discrimination: a counter-evaluation. *Perception*, 14, 471-497.

Vickers, D. and Smith, P.L. (1986). The rationale for the inspection time index. *Personality and Individual Differences*, 7, 609-624.

Warburton, D.M. (1991). From molecules to cognitions. *The Psychologist*, 4, 533-536.

Warren, R.M., Obusek, C.J., Farmer, R.M. and Warren, R.P. (1969). Auditory sequence: confusion of pattern other than speech or music. *Science*, 164, 586-587.

Warren, R.M. and Warren, R.P. (1970). Auditory illusions and confusions. *Scientific American*, 223, December, 30-36.

Watson, B.U. (1991). Some relationships between intelligence and auditory processing. *Journal of Speech and Hearing Research*, 34, 621-627.

Watson, C.S., Johnson, D.M., Lehman, J.R., Kelly, W.J. and Jensen, J.K. (1982). An auditory discrimination test battery. *Journal of the Acoustical Society of America*, 71 Suppl. 1, S73.

Watson, R.I. (1979). *R.I. Watson's Selected Papers on the History of Psychology*. Edited by J. Brozek and R.B. Evans. University of New Hampshire.

Weinberg, H. (1969). Correlation of frequency spectra of averaged visual evoked potentials with verbal intelligence. *Nature*, 224, 813-814.

Weiss, V. (1984). Psychometric intelligence correlates with interindividual differences in rates of lipid peroxidation. *Biochemica and Biophysica Acta*, 43, 755-763.

Weiss, V. (1987). The quantum mechanics of EEG-brain dynamics and short term memory. *Biol. Zent. Bl.*, 106, 401-408.

Whitely, S.E. (1980). Latent trait models in the study of intelligence. *Intelligence*, 4, 97-132.

Widaman, K.F. and Carlson, J.S. (1986). Procedural effects on performance on the Hick paradigm. *Intelligence*, 13, 63-86.

Wilson, J.T.L., Wiedmann, K.D., Phillips, W.A. and Brooks, D.N. (1988). Visual event perception in alcoholics. *Journal of Clinical and Experimental Neuropsychology*, 10, 222-234.

Wissler, C. (1901). The correlation of mental and physical tests. *Psychological Review*, Monograph No. 3.

Wuketits, F.M.P. (1986). Evolution as a cognition process: towards an evolutionary epistemology. *Biology and Philosophy*, 1, 191-206.

Zhang, Y., Caryl, P.G. and Deary, I.J. (1989a). Evoked potential correlates of inspection time. *Personality and Individual Differences*, 10, 379-384.

Zhang, Y., Caryl, P.G. and Deary, I.J. (1989b). Evoked potentials, inspection time and intelligence. *Personality and Individual Differences*, 10, 1079-1094.

Zohary, E., Hillman, P. and Hochstein, S. (1990). Time course of perceptual discrimination and single neuron reliability. *Biological Cybernetics*, 62, 475-486.

**Appendix 1.1**

Ability test correlation matrix for boys (n=43) from the study by Abelson (1911).

	1	2	3	4	5	6
1 - Reading ability	-	0.54	0.20	0.20	0.40	0.11
2 - Practical intelligence		-	0.18	0.32	0.28	0.47
3 - Memory for sentences			-	0.39	0.21	0.17
4 - Geometrical figures				-	0.17	0.21
5 - Crossing out dots					-	0.46
6 - Discrimination of lines						-
	7	8	9	10	11	12
1 - Reading ability	0.19	0.74	0.41	0.37	0.20	0.41
2 - Practical intelligence	0.52	0.46	0.24	0.28	0.19	0.04
3 - Memory for sentences	0.36	0.41	0.38	0.20	0.66	0.02
4 - Geometrical figures	0.36	0.30	0.56	0.07	0.52	0.07
5 - Crossing out dots	0.15	0.39	0.04	0.32	0.06	0.65
6 - Discrimination of lines	0.21	0.20	0.34	0.38	0.18	0.31
7 - Interpreting pictures	-	0.14	0.22	0.25	0.30	0.27
8 - Arithmetical ability		-	0.10	-0.11	0.32	-0.02
9 - Memory for commissions			-	0.38	0.38	-0.15
10 - Tapping				-	-0.08	0.46
11 - Memory for object names					-	-0.22
12 - Crossing out rings						-



**Appendix 1.2**

Ability test correlation matrix for girls (n=88) from the study by Abelson (1911).

	1	2	3	4	5	6
1 - Practical intelligence	-	0.52	0.42	0.43	0.45	0.39
2 - Memory for commissions		-	0.14	0.30	0.50	0.30
3 - Tapping			-	0.42	0.27	0.61
4 - Crossing out rings				-	0.21	0.26
5 - Memory for sentences					-	0.49
6 - Interpreting pictures						-
	7	8	9	10	11	12
1 - Practical intelligence	0.51	0.43	0.35	0.33	0.18	0.43
2 - Memory for commissions	0.41	0.31	0.43	0.31	0.34	0.37
3 - Tapping	0.29	0.33	0.37	0.21	0.30	0.26
4 - Crossing out rings	0.34	0.32	0.42	0.47	0.17	0.27
5 - Memory for sentences	0.16	0.43	0.32	0.26	0.42	0.13
6 - Interpreting pictures	0.30	0.42	0.26	0.25	0.33	0.00
7 - Arithmetical ability	-	0.32	0.26	0.30	0.30	0.47
8 - Geometrical figures		-	0.21	0.45	0.13	0.01
9 - Discrimination of lines			-	0.22	0.22	0.20
10 - Crossing out dots				-	0.18	0.15
11 - Memory for object names					-	0.24
12 - Reading ability						-

### Appendix 1.3

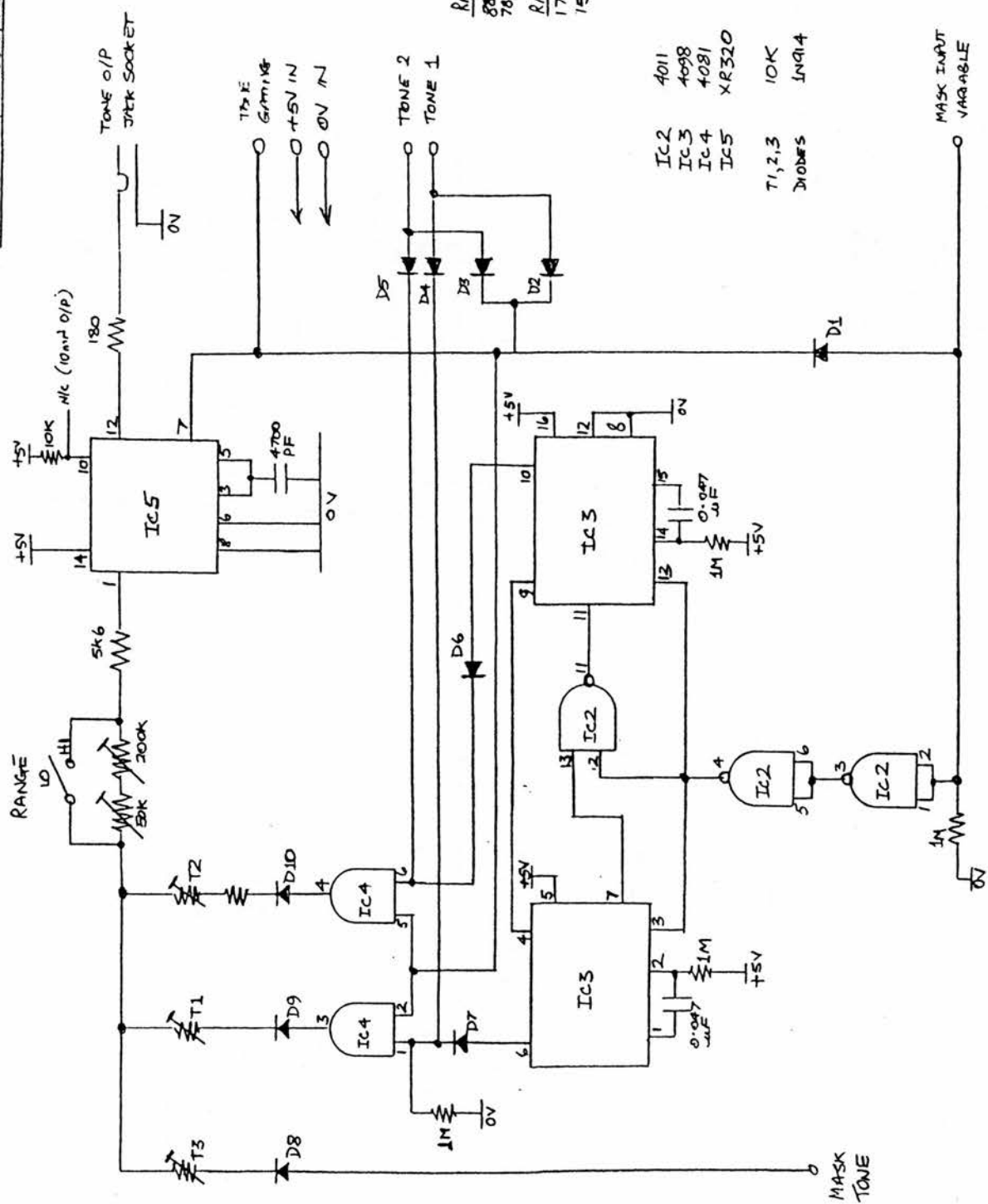
Ability test correlation matrix for all subjects (n>150) from the studies by Carey (1915-17). Coefficients in the diagonal are test-retest reliabilities.

	1	2	3	4	5	6	7	8
1 - School intelligence	0.79	0.64	0.64	0.56	0.34	0.24	0.14	0.27
2 - Practical intelligence		0.60	0.33	0.19	0.29	0.19	-0.09	0.15
3 - Painstaking			0.62	0.60	0.31	0.15	0.14	0.23
4 - Social status				0.72	0.20	0.28	0.05	0.05
5 - Visual discrimination					0.53	0.26	0.00	0.45
6 - Auditory discrimination						0.72	0.01	0.15
7 - Tactile discrimination							0.57	0.09
8 - Visual memory								0.64
		9	10	11	12	13	14	15
1 - School intelligence		0.38	0.38	0.82	0.63	0.49	0.35	0.26
2 - Practical intelligence		0.19	0.34	0.59	0.42	0.36	0.20	0.23
3 - Painstaking		0.26	0.40	0.53	0.36	0.29	0.41	0.21
4 - Social status		0.26	0.10	0.18	0.28	0.18	0.24	0.32
5 - Visual discrimination		0.38	0.22	0.32	0.51	0.17	0.33	0.14
6 - Auditory discrimination		0.37	-0.01	0.23	0.46	0.27	0.19	0.23
7 - Tactile discrimination		0.11	-0.11	0.09	-0.05	-0.04	-	-
8 - Visual memory		0.44	0.28	0.18	0.42	0.31	0.25	0.03
9 - Auditory memory		0.68	0.33	0.09	0.46	0.22	0.23	0.05
10 - Verbal memory			0.81	0.25	0.31	0.34	0.17	0.03
11 - Exam marks				0.86	0.45	0.62	0.54	0.10
12 - Opposites					0.79	0.70	0.23	0.14
13 - Sentences						0.76	0.24	0.15
14 - Drawings							0.72	0.51
15 - Holes								0.74

## **Appendix 2**

Overleaf is the circuit diagram for the device which generated the auditory stimuli used to estimate auditory inspection times in Chapters 4, 5, 7 and 8. The device was designed and constructed by the late Mr D. Wight, sometime Senior Electronics Technician, Department of Psychology, University of Edinburgh. The diagram was drawn by Mr G. Baldwin, Senior Electronics Technician, Department of Psychology, University of Edinburgh.

# GENERATOR AND CONTROL BOARD



RANGE	LO	HI	LO
	880 Hz	784 "	LO
RANGE	HI	HI	LO
	1760 Hz	1588 Hz	LO

IC2 4011  
IC3 4098  
IC4 4081  
IC5 XR320  
T1,2,3 10K  
DIODES 1N914

### Appendix 3

Results of Probit analyses of the auditory inspection time data for all subjects in Chapter 5 who supplied full data, divided by whether or not they were classified as being able to perform the AIT task.

#### Subjects from study 1 in Chapter 5 who could do AIT

ID No.	Regress. Coeff.	S.E.	Coeff./ S.E.	Chi Squ.	p value	90% Thr	50% Thr	90% CI
B01	1.753	.329	5.328	35.029	.000	72.1	13.4	1.0-27.3
B02	1.625	.294	5.527	36.542	.000	147.9	24.1	4.1-50.0
B04	1.505	.303	4.967	22.771	.019	95.7	13.5	2.1-25.8
B05	1.592	.329	4.838	18.510	.070	64.4	10.1	1.8-18.9
B08	1.738	.316	5.500	14.761	.194	87.7	16.1	9.1-23.1
B10	1.260	.286	4.406	22.070	.024	125.6	12.1	0.7-25.6
B13	1.675	.335	4.717	15.011	.182	50.1	8.6	3.4-13.8
B20	2.705	.565	4.789	10.904	.451	29.1	9.8	5.2-13.3
B22	1.361	.305	4.465	16.564	.121	84.2	9.6	1.3-19.0
B23	1.041	.267	3.901	44.153	.000	267.9	15.8	-
B24	1.859	.346	5.376	11.911	.370	62.0	12.7	6.9-18.4
B27	1.499	.352	4.256	22.316	.022	46.7	6.5	0.1-14.9
B28	2.036	.404	5.034	5.354	.913	41.7	9.8	5.0-14.4
B34	1.351	.287	4.711	26.719	.005	131.5	14.8	1.0-31.0
B43	1.339	.301	4.457	14.923	.186	91.1	10.1	3.4-17.0
B44	1.810	.369	4.902	9.756	.552	47.6	9.3	4.2-14.3
B50	1.798	.311	5.781	20.211	.043	106.4	20.6	8.6-33.8
C02	2.141	.378	5.662	8.953	.626	54.4	13.7	8.3-19.0
C04	1.392	.291	4.786	21.669	.027	119.9	14.4	2.1-27.9
C05	2.137	.437	4.893	9.827	.546	36.5	9.2	4.6-13.4
C06	0.917	.263	3.485	15.388	.165	338.0	13.5	2.7-25.4
C17	1.389	.292	4.758	11.740	.383	115.8	13.8	6.1-21.8
C22	1.207	.293	4.114	19.026	.061	101.9	8.8	0.3-19.7
C24	1.224	.272	4.504	21.898	.025	240.0	21.6	3.5-43.3
C27	0.819	.270	3.035	23.283	.016	239.1	6.5	-
C30	1.200	.274	4.380	16.478	.124	199.6	17.1	3.4-32.2
C31	0.584	.274	2.130	19.909	.047	279.5	1.8	-
C32	1.669	.339	4.918	6.633	.828	59.2	10.1	4.5-15.7
C34	0.523	.249	2.100	25.635	.007	4365.1	15.4	-
C37	1.711	.333	5.144	13.057	.290	66.2	11.8	5.8-17.8
C38	1.620	.388	4.173	14.236	.220	36.6	5.9	1.6-10.5
C40	1.389	.280	4.955	23.102	.017	185.8	22.2	5.2-42.2
C41	2.128	.404	5.270	15.954	.143	44.2	11.1	4.3-17.6
C42	1.270	.287	4.423	15.805	.149	123.7	12.1	2.0-23.2
C45	1.182	.280	4.225	25.682	.007	151.2	12.5	0.1-28.6
C48	1.002	.271	3.698	27.321	.004	210.6	11.7	0.0-29.8
D05	0.962	.272	3.541	8.946	.627	204.3	9.5	1.4-18.9
D10	1.653	.370	4.464	14.816	.191	43.0	7.2	2.4-12.1
D15	1.253	.297	4.213	10.598	.478	94.0	8.9	2.5-15.9
D20	2.026	.385	5.270	12.031	.361	47.8	11.2	6.1-16.1
D21	1.870	.344	5.430	10.614	.476	63.3	13.1	7.3-19.0
D22	0.875	.272	3.220	13.363	.270	211.4	7.3	0.5-16.3
D24	1.347	.292	4.606	5.888	.881	111.0	12.4	4.9-20.1
D26	1.375	.308	4.456	14.172	.224	79.2	9.3	3.1-15.7
D34	1.739	.332	5.237	14.409	.211	67.9	12.4	6.4-18.5
D35	0.862	.263	3.279	12.004	.363	344.4	11.2	1.4-22.7
D37	1.123	.277	4.054	17.365	.098	164.9	11.9	0.81-25.0
D39	2.477	.644	3.846	8.949	.627	21.0	6.4	2.4-9.7

D43	1.140	.276	4.123	17.974	.082	169.6	12.7	1.0-26.5
	<b>Regress.</b>		<b>Coeff./</b>					
<b>ID No.</b>	<b>Coeff.</b>	<b>S.E.</b>	<b>S.E.</b>	<b>Chi Squ.</b>	<b>p value</b>	<b>90 %Thr</b>	<b>50 %Thr</b>	<b>90 %CI</b>
D46	1.568	.310	5.066	13.811	.244	87.0	13.25	6.4-20.2
D47	1.853	.331	5.596	36.444	.000	74.9	15.3	1.9-30.2

### Subjects from study 1 in Chapter 5 who couldn't do AIT

<b>ID No.</b>	<b>Regress.</b>		<b>Coeff./</b>					
	<b>Coeff.</b>	<b>S.E.</b>	<b>S.E.</b>	<b>Chi Squ.</b>	<b>p value</b>	<b>90 %Thr</b>	<b>50 %Thr</b>	<b>90 %CI</b>
B03	0.517	.249	2.076	18.192	.077	3867	12.8	-
B06	0.349	.248	1.407	13.773	.246	>10000	6.28	-
B07	0.139	.246	0.565	13.627	.254	>1000000	0.5	-
B09	0.145	.244	0.594	13.019	.292	>1000000	9.65	-
B19	1.116	.266	4.201	11.549	.399	529.6	37.6	20.9-84.6
B38	0.702	.255	2.756	17.712	.089	892.2	13.3	-
B39	0.075	.243	0.312	4.584	.950	>1000000	144.6	-
B42	-0.007	.244	-0.027	17.498	.094	<0.0000001	>1000000	-
B46	-0.036	.244	-0.146	6.038	.871	<0.0000001	>10000	-
B47	0.383	.247	1.551	15.301	.169	>10000	80.4	-
B49	-0.415	.246	-1.686	15.590	.157	0.02	25.9	-
C03	0.043	.244	0.174	14.219	.221	>1000000	>10000	-
C07	0.617	.252	2.454	11.676	.388	1706	14.3	0.1-33.5
C08	0.260	.248	1.048	18.767	.065	>100000	2.3	-
C09	0.417	.247	1.691	11.639	.391	>10000	16.9	-
C18	0.231	.244	0.946	15.247	.171	>1000000	11.5	-
C21	0.725	.252	2.871	14.942	.185	2170	37.0	11.7-84.0
C26	-0.494	.248	-1.995	17.857	.085	0.05	21.4	-
C28	0.125	.244	0.519	10.044	.526	>1000000	3.7	-
C33	0.247	.244	1.009	14.101	.227	>1000000	15.0	-
C49	0.302	.248	1.234	8.335	.683	>100000	28.7	-
C50	0.310	.246	1.260	22.543	.020	>100000	10.4	-
D01	0.065	.244	.265	16.585	.121	>1000000	0.2	-
D04	0.310	.247	1.254	14.601	.202	>10000	5.0	-
D09	0.160	.244	0.658	16.011	.141	>1000000	14.7	-
D11	0.429	.249	1.724	7.837	.728	7771	8.1	-
D19	0.339	.245	1.384	8.461	.672	>100000	34.4	-
D25	0.499	.248	2.017	20.498	.039	8668	23.5	-
D28	0.501	.253	1.980	7.640	.745	2274	6.3	0.0-21.4
D32	0.692	.252	2.74	11.470	.405	3455	48.5	18.1-147.5
D38	0.334	.246	1.359	11.698	.387	>10000	13.3	-
D41	0.920	.260	3.536	15.491	.161	447.4	18.1	4.9-32.4
D42	0.575	.253	2.275	13.800	.244	1606	9.5	0.0-25.6



**Appendix 3 (continued)**

Cognitive ability test scores of subjects included in Appendix 3.

<b>Identity No.</b>	<b>Alice Heim 6 Verbal</b>	<b>Alice Heim 6 Numerical &amp; Diagrammatic</b>	<b>Alice Heim 6 Total</b>
B01	12	12	24
B02	15	15	30
B03	15	10	25
B04	17	18	35
B05	26	24	50
B06	21	20	41
B07	17	13	30
B08	15	10	25
B09	9	17	26
B10	20	17	37
B13	17	10	27
B19	17	6	23
B20	27	21	48
B22	11	17	28
B23	22	22	44
B24	15	15	30
B27	20	17	37
B28	15	14	29
B34	12	14	26
B38	11	10	21
B39	18	20	38
B42	15	11	26
B43	16	11	27
B44	17	16	33
B46	17	11	28
B47	27	20	47
B49	17	12	29
B50	19	15	34
C02	16	12	28
C03	15	9	24
C04	14	8	22
C05	14	17	31
C06	13	12	25
C07	22	22	44
C08	17	14	31
C09	15	9	24
C17	14	13	27
C18	9	12	21
C21	14	13	27
C22	18	16	34
C24	15	11	26
C26	22	15	37
C27	16	9	25
C28	7	10	17
C30	14	13	27
C31	15	9	24
C32	17	17	34
C33	17	17	34
C34	14	10	24
C37	15	9	24
C38	12	12	24

<b>Identity No.</b>	<b>Alice Heim 6 Verbal</b>	<b>Alice Heim 6 Numerical &amp; Diagrammatic</b>	<b>Alice Heim 6 Total</b>
C40	10	9	19
C41	16	16	32
C42	21	17	38
C45	10	11	21
C48	11	15	26
C49	12	14	26
C50	19	20	39
D01	16	11	27
D04	14	13	27
D05	11	10	21
D09	7	4	11
D10	20	14	34
D11	7	22	29
D15	11	10	21
D19	16	16	32
D20	16	19	35
D21	16	15	31
D22	16	15	31
D24	15	5	20
D25	22	16	38
D26	16	15	31
D28	9	15	24
D32	9	10	19
D34	13	13	26
D35	11	11	22
D37	24	24	48
D38	15	14	29
D39	16	11	27
D41	16	15	31
D42	15	11	26
D43	7	8	15
D46	22	22	44
D47	14	14	28

## Appendix 4

Formula for calculating the  $\bar{Z}_2^*$  statistic for comparing the size of two dependent correlations, in this case comparing the correlation between variables j and k with the correlation between variables h and m (Steiger, 1980).

$$\bar{Z}_2^* = (N - 3)^{0.5} (z_{jk} - z_{hm}) (2 - 2\bar{s}_{jk,hm})^{-0.5} \quad (1)$$

where,

N = number of subjects,

z = Fisher's transformation of Pearson's  $r$  between the variables of interest,

j, k, h, and m = the four variables in the correlation matrix,

and where,  $\bar{s}_{jk,hm}$  = a special case of the equation,

$$c_{jk,hm} = \psi_{jk,hm} / (1 - \rho_{jk}^2) (1 - \rho_{hm}^2) \quad (2)$$

where pooled estimates replace the sample correlations  $\rho_{jk}$  and  $\rho_{hm}$

and where,

$$\begin{aligned} \psi_{jk,hm} = 0.5 \{ & [(\rho_{jh} - \rho_{jk}\rho_{kh}) X (\rho_{km} - \rho_{kh}\rho_{hm})] + \\ & [(\rho_{jm} - \rho_{jh}\rho_{hm}) X (\rho_{kh} - \rho_{kj}\rho_{jh})] + \\ & [(\rho_{jh} - \rho_{jm}\rho_{mh}) X (\rho_{km} - \rho_{kj}\rho_{jm})] + \\ & [(\rho_{jm} - \rho_{jk}\rho_{km}) X (\rho_{kh} - \rho_{km}\rho_{mh})] \} \end{aligned} \quad (3)$$

## Appendix 5

Published papers. The following papers appear overleaf. I am grateful to Peter Caryl, Vincent Egan and Brian Head for their permission to include them in the thesis.

1. Deary, I.J. (1988). Prospects for the biology of intelligence. *Human Evolution*, 3, 503-513.
2. Deary, I.J., Caryl, P.G., Egan, V. and Wight, D. (1989). Visual and auditory inspection time: their interrelationship and correlations with IQ in high ability subjects. *Personality and Individual Differences*, 10, 525-533.
3. Deary, I., Head, B. and Egan, V. (1989). Auditory inspection time, intelligence and pitch discrimination. *Intelligence*, 13, 135-147.
4. Deary, I.J. (1989). Auditory inspection time: review, recent research and reassessment. In D. Vickers and P.L. Smith (Eds), *Human Information Processing: Measures, Mechanisms and Models*. Amsterdam: Elsevier.
5. Deary, I.J. and Caryl, P.G. (1988). Inspection time and cognitive ability. In K.J. Gilhooly, M.T.G. Keane, R.H. Logie and G. Erdos (Eds), *Lines of Thinking: Reflections on the Psychology of Thought. Volume 2: Skills, Emotion, Creative Processes, Individual Differences and Teaching Thinking*. Chichester: Wiley.

**I.J. Deary**

*Department of Psychology  
University of Edinburgh  
7, George Square  
Edinburgh EH8 9JZ  
Scotland*

**Prospects for the Biology of Human Intelligence**

Despite the ubiquitous nature of Spearman's *g* in mental test performance, the charge «intelligence is what intelligence tests test» has not been countered in a satisfactory way. It is proposed that there are two ways to answer this complaint. The first concerns the new hypothesis testing models in factor analysis. The second involves studying the 'biology of intelligence'. The biology of intelligence has various meanings and four are discussed: biology as theory; biology as race and genetics; biology as neurobiology; and biology as basic psychological processes. The last of these is considered in some detail and it is found that reaction time, evoked potentials and inspection time offer bright prospects for further research on the biology of psychometric intelligence.

*Key words:* Intelligence, inspection time, reaction time, evoked potentials

**Introduction**

Tests of intelligence are unpopular with Western intellectuals today. Notoriously, the tests appear to discriminate against «minorities» and against «the working class»; and women are under-represented in the higher ranges of IQ. Moreover, the tests are often held to be «circular»: critics who make this charge are repeating, however unknowingly, the dictum of one of the few geniuses who have ever applied their talents to psychology — namely Edwin Boring, who was the first to remark, in 1923, that intelligence «is what the tests test».

Until recently it has not been possible to answer this objection definitively, even if one accepts the major discovery of the psychometricians that, given a large random sample of the population, when a group is tested on a variety of different mental tests the correlation matrix that results is almost entirely positive. This is the finding that led SPEARMAN (1904) to propose that, due to differences in brain functioning, people had reliable individual differences in general intelligence (or *g*). This conclusion has not proved easy to substantiate. For, while Spearman's positive manifold is a common finding, its interpretation is not unproblematic.

A recent principal components analysis of the 11 very different subtests of the WAIS-R (CANAVAN, DUNN & McMILLAN, 1986) yielded a *g* factor which accounted for over 55% of the variance between subjects on the tests. In this study no subtest had a loading of less than 0.64 on the first, general factor. More generally, attempts to replace conventional IQ-style tests with non-*g* loaded mental ability tests have not been successful. HOOPER, HOOPER & COLBERT (1984) reported a study involving subjects of different age groups who were administered standard psychometric intelligence tests alongside tests developed from PIAGET's theory of formal operational thinking. The average correlation between Raven's Matrices and twelve formal operational reasoning tasks was 0.53. These two studies are merely illustrative of many similar efforts. These are typical findings and demonstrate the ubiquitous nature of *g*.

The problem of the interpretation of *g* lies in the reliance on specific statistical methodology. It is possible to choose a particular factor-analytic method and, from the

characteristics of the method, fashion one's own model of intelligence. Thus, followers of Spearman such as BURT (1909-10); JENSEN (1980) and EYSENCK (1982) have continued to extract the general factor using methods like principal components analysis. Others claim that this method tends to ensure a general factor. Loudest among these remains GUILFORD (1985) who, Nelson-like, has not «seen any g» because he has chosen to implement factor analytic techniques that ignore g. Thus he has been able to construct a model of intelligence which contains 120 individual abilities although, as EYSENCK (1979) pointed out, many of these are correlated and, if submitted to higher order analysis, will yield fewer and more general ability factors.

How can such a strong empirical finding — the positive manifold of mental test correlation — give rise to such very different models? BORING (1923) had a point. Within the psychometric field there has been no resolution of the various general or multi-box models of intelligence. There is a need to find some external criterion or some antecedent variable that is related to performance on mental ability tests. Resolution may be coming in two ways.

First, the nature of factor analysis is changing. It is now possible actively to test hypothetical models of intelligence with techniques such as LISREL. Using this method GUSTAFSSON (1984) administered 16 tests of intelligence to 1000 subjects and tested various models of intelligence for their goodness of fit to the data. Interestingly, he was able to give comfort to many previous theorists. Older models of the structure of human abilities appeared to be special cases of his hierarchical unifying model of intelligence. At the lowest level in the hierarchy of ability lay primary factors similar to those proposed by THURSTONE (1938) and GUILFORD (1985). At a higher level there appeared fluid and crystallised intelligence, as hypothesised by CATTELL (1963). At the peak of the hierarchy lay g, which was indistinguishable from the second order fluid intelligence factor. Models which had seemed irreconcilable for so long emerged as complementary rather than contradictory; in the past various authors have merely drawn attention to various parts of the larger model.

### **The Biology of Intelligence**

Second, and more important, escape from circularity has come with the advent of an area of study known as the «biology of intelligence». But this term has its problems: it immediately alienates those who see this as an exercise in genetic reductionism in what appears to be an arbitrary and culture-biased set of puzzles (GOULD, 1981; ROSE, LEWONTIN & KAMIN, 1984). To others the term is problematic because of its diversity, for it applies to no one area of study and to no one set of methods. In what follows I will examine the experimental evidence for one aspect of the biology of intelligence — namely, those «basic» psychological processes that correlate with IQ test scores. But, before that, it is useful to catalogue the various approaches that have been taken by other researchers.

First, biology as a model. In ROBERT STERNBERG's (1985) review of models of intelligence he included PIAGET's work as an example of the biological approach. In this sense Piaget's contribution has been important and very different from that of the psychometricians. While PIAGET's tests of conservation, etc. have provided mental tests with moderate g-loadings (and much variance that is specific to the tests themselves (see JENSEN, 1980)), his work on the biology of knowledge growth has been much less studied. Knowledge accretion, according to PIAGET (1971, 1978, 1980) involves the brain acting in the way that other organs do: with a substrate (information from the world) and with products which are transforms of the substrate (schemata, formed by the processes of

assimilation and accommodation). PIAGET's writings on the growth of intelligence have a structure that is close to that of the evolutionary epistemologists (WUKETITS, 1986; CAMPBELL, 1974) and those AI workers whose first assumptions and constraints involve what is known about brain development (EDELMAN & REEKE, 1982). This form of the biology of knowledge might form a superstructure for psychometricians but it is not of immediate empirical concern.

Second, biology as race, genetics and heredity. It is this aspect of the psychometric endeavour that has coloured many others and has often prevented rational argument. The discovery of the Burt fraud (HEARNshaw, 1979) and the reactions to the work of JENSEN (1969) and EYSENCK (1973) by non-specialist writers like KAMIN (1974); GOULD (1981) and ROSE, LEWONTIN & KAMIN (1984) have made others wary of the field. JENSEN argued recently (1985a; there are many peer criticisms appended to the article) that the one standard deviation difference between US blacks and whites on IQ test scores has its basis in the tests' g-loadings. Nevertheless, JENSEN also reports that blacks are superior to whites on other abilities such as memory. On the other hand, MACKINTOSH's (1986) research in this area has drawn attention to evidence that fails to support a genetic origin for mental ability differences between ethnic groups: MACKINTOSH implicates differences in the social circumstances of the different ethnic groups.

Third, biology as neurobiology. This endeavour involves searching for biological, often biochemical, correlates of intelligence test scores. Patient groups with impaired levels of cognition such as Alzheimer or Down's syndrome patients are often used. WEISS (1984, 1986) has reviewed evidence that Down's syndrome involves excess peroxidation of neuronal membranes and a build-up of the products of oxidation within the neuron. Both of these, he claims, will impair information transfer. One enzyme involved in the prevention of excess oxidation is glutathione peroxidase (GSHPx). SINET, LEJEUNE & JEROME (1979) argued that the level of this enzyme in Down's syndrome patients would be correlated with the level of accurate information transfer and that this in turn would be demonstrated by a correlation between the enzyme level and the patient's IQ. They reported a correlation of 0.5 between GSHPx with IQ scores in 50 Down's syndrome patients. Uric acid is also involved in the prevention of lipid oxidation by free radicals and INOUE, PARK & ASAKA (1984) have reported a correlation of 0.334 ( $p < .025$ ) between uric acid level and IQ. These workers have argued, from twin data, that the two traits (uric acid level and IQ) might have partial communality of gene loci.

Using Alzheimer patients and normal controls CHASE, FEDIO, FOSTER, BROOKS, DiCHIRO & MANSI (1984) found a correlation of 0.68 between glucose use (as measured by positron emission tomography scanning of fluorodeoxyglucose F18) in the cerebral cortex and full scale IQ on the WAIS. Further evidence for the biological basis of IQ scores came from their localisation data: verbal IQ related best to glucose metabolism rate in the left temporal region ( $r = 0.76$ ) and performance IQ correlated best with glucose metabolism in the right parietal region ( $r = 0.70$ ). Also using Alzheimer patients, DEARY, HENDRICKSON & BURNS (1987) have demonstrated a correlation of 0.5 ( $p < .02$ ) between serum calcium levels and cognitive scores assessed by the Mini Mental State (FOLSTEIN, FOLSTEIN & McHUGH, 1975). These authors have argued that calcium has a central role in the maintenance of the integrity of the normal cytoskeleton and that low calcium levels impair information transfer.

### Basic Processes in Intelligence

Fourth, biology as basic psychological processes. In this sense biology refers to psychological tests that index purported basic information-processing functions and con-



straints. The renewed interest in basic processes marks a return to the empirical efforts of GALTON (1883), SPEARMAN (1904) and Spearman's students such as ABELSON (1911) and CAREY (1914-15). The success of the Binet approach to intelligence testing — the estimation of general ability using a hotch-potch of mental tests with appropriate weighting before an average is taken — has had its drawbacks. In 1911 Abelson warned that the immediate practicality of the Binet test, (in education, mental handicap, job selection, etc.) which measured «higher-level» processes like memory, reasoning and judgement, meant that psychometry would become divorced from experimental psychology. The worry was that the more the tests proliferated the less was becoming known about the psychological processes that underlie success in performing them. Although there has been a trickle of experimental studies hinting that perceptual speed, reaction time and sensory discrimination were reliable and moderately high correlates of IQ test performance (see DEARY, 1986 and 1987 for reviews), until the last decade there has been no concerted research effort to discover what amount, if any, of the interindividual variance on IQ scores is attributable to individual differences in simpler psychological performance.

This revived effort has been driven by two ideas. First, the idea that the more intelligent person has some advantage in mental speed has always had some currency. Experimental evidence (see below), professional and lay opinions agree that intelligent people tend to be «quick-witted», «quick on the uptake» and «quick thinkers» (STERNBERG, CONWAY, KETRON & BERNSTEIN, 1981). Second, the researchers in this field have implicitly or explicitly accepted that the solution of IQ test items involves many psychological processes and that these may be expressed as a hierarchy. Thus Gustafsson's hierarchy of psychometric intelligence is mirrored by an internal hierarchy of psychobiological processes. This hypothesis proposes that (JENSEN, 1985b) IQ test items are solved by psychological metaprocesses. These metaprocesses are combinations of basic psychological processes and their orchestration and combination are affected by prior experience, education and coaching. Thus the efficiency of metaprocesses will have some imperfect correlation with the functioning of individual basic processes. JENSEN (1985b) suggests that these basic processes might include stimulus apprehension, iconic memory, stimulus encoding, short term memory, rehearsal of short term memory, memory scanning, retrieval from long term memory, mental rotation, response execution, etc. At an even more basic level these processes will share the performance constraints of a common neurology and thus their efficiency may be correlated to yield a biological general intelligence factor. Thus the efficiency of neuronal transmission (affected by inherited factors and environmental factors such as nutrition and exposure to neurotoxic agents) may correlate with the basic psychological processes (affected by time of day, sedative drugs, etc.) which combine to form metaprocesses which solve informational problems. These problem solutions are factor analysed to yield ability clusters (verbal, visuo-spatial, etc) which correlate and yield a psychometric g factor. Therefore both biological g and psychometric g are hypothetical and not, as yet, able to be indexed directly.

Three more or less basic psychological tests have attracted much interest in the last decade as correlates of intelligence. All involve some form of mental speed and are: reaction time (RT); average evoked potentials (AEPs); and inspection time (IT).

### **Intelligence and Reaction Time**

Reaction time measures have a history of reliable but modest correlation with IQ scores (see BECK, 1933 for a review of the early studies) but the current research has been dominated by the work of Jensen using the HICK (1952) paradigm. In the Hick method

subjects are positioned before a response panel and place their preferred index finger on a «home» button. Surrounding the home button is a semicircle of 8 equidistant lights, each of which has a response button in front of it. The task involves waiting for a light to come on whereupon the subject must release the home button and, as quickly as possible, move to the button in front of the appropriate light and press it. This yields two measures: the time it takes for the subject to release the home button after the light has been switched on is termed the response time (RT); and the time from the release of the home button until the button in front of the target light is depressed is called the movement time (MT). Together they are called the reaction time. The Hick paradigm involves varying the number of lights from which the target light may be expected: Jensen uses 1, 2, 4 and 8 light sets. Thus reaction times to 0, 1, 2 and 3 bits of information may be assessed. Hick's law states that there is a linear increase in reaction time as the number of bits of information increases (i.e.  $\log_2$  of the number of stimulus alternatives).

JENSEN (1986) has collected 20 independent subject samples where the slope of Hick RT has been correlated with mental ability test scores. Twelve of the studies are from Jensen's laboratory (1055 subjects): of 22 correlations the N-weighted mean  $r$  is  $-0.091$ ,  $SD = 0.109$ . Eight studies from other laboratories (503 subjects) have delivered 13 correlations: their N-weighted mean is  $-0.181$ ,  $SD = 0.147$ . Jensen makes two important points about these correlations. First, despite their low correlation the significance of these results is at the 0.1% level. Second, many of the samples include university students whose ability range is less than half of the population range: this reduces the size of the correlations and Jensen estimates that a true population correlation between Hick RT slope and IQ would be in the region of  $-0.3$ .

JENSEN (1986) has also reviewed the work on the correlation between IQ and the RT for the individual stimulus set sizes on the HICK paradigm. From 31 studies (1129 subjects) the N-weighted mean correlations are:  $-0.18$  for 1 light (0 bits);  $-0.19$  for 2 lights (1 bit);  $-0.22$  for 4 lights (2 bits); and  $-0.23$  for 8 lights (3 bits). Again, 19 of these studies are from Jensen's own laboratory. Two recent critical accounts of this work have appeared. LONGSTRETH (1986) proposes that the large number of studies that come from the same laboratory or from the laboratories of Jensen's former students (especially P.A. Vernon) biases the results in some unspecified way. He also suspects that negative studies are less likely to be published hence elevating the reported mean correlation. MACKINTOSH (1986) has suggested that RT is not a basic process: he attributes the correlations to subjects' willingness to concentrate on a boring task. FREARSON & EYSENCK (1986), using 37 normal adults tested on Raven's APM, found an IQ-RT correlation of about  $-0.3$  and an IQ-MT correlation of  $-0.45$  regardless of the number of stimulus alternatives, contrary to Jensen's findings. When these authors made the Jensen task more cognitively demanding by asking subjects to respond to one of three lights on the basis of its relative position the IQ-RT correlation increased to between  $-0.5$  and  $-0.6$ . These results indicate that speed of RT and MT to even a single stimulus is significantly correlated to IQ but, obviously, if the task is made more difficult, the correlation will rise (because, presumably, extra basic psychological processes are being sampled).

## Intelligence and Evoked Potentials

The relationship between IQ scores and indices derived from AEPs to simple auditory stimuli has been investigated most recently by A.E. & D.E. HENDRICKSON (1980, 1982). Earlier work by SCHUCARD & HORN (1972) and ERTL & SCHAFFER (1969) had

demonstrated a low correlation between amplitude and latency measures of AEPs and psychometric intelligence. The Hendricksons have used the «string length» of the AEP, a measure of AEP complexity. They argue that high IQ subjects are characterised by few errors in neuronal transmission while low IQ subjects make more errors on average. Thus, they claim, when a large number of evoked potentials to an identical stimulus are averaged for a high IQ subject the low error rate (over many neurons) will result in similar waveforms. For the low IQ subject the higher error rate will result in more variability in individual evoked potentials. Therefore the low IQ AEP loses much of the complexity of the individual evoked potentials that compose it. The Hendricksons measure complexity by a computer-run algorithm which, in essence, places a string over the AEP for a given epoch and measures its length.

Using the average of 90 EPs to 85 dB, 30 ms, 1000 Hz tones BLINKHORN & D.E. HENDRICKSON (1982) correlated «string» measures with Raven's Advanced Progressive Matrices Scores on 34 students (17 male, 17 female). Post stimulus 512 ms strings correlated 0.538 ( $p < .001$ ) with APM scores. Unstimulated EEG strings correlated at 0.127 (n.s.) with APM scores. HENDRICKSON & HENDRICKSON (1982) reported a correlation of 0.72 in 219 schoolchildren. HAIER, ROBINSON, BRADEN & WILLIAMS' (1983) replication found correlations between 0.13 and 0.50 (many were non-significant) although they did find that stimulus intensity was important. In a small study where the Hendricksons' methods were followed closely, CARYL & FRASER (1985) found a correlation of 0.72 between Alice Heim test scores and string length measures of AEP. MACKINTOSH (1986) failed to replicate this finding and suggests, again, that while the work of the Hendricksons is important, their results might be due to a willingness to comply with and to continue to concentrate upon a boring task.

### Intelligence and Inspection Time

Inspection time is the newest of the three basic measures. Although similar tasks were shown to be useful in discriminating subjects of different levels of ability many decades ago (CATTELL, 1886; BURT 1909-10), the IT task as it is used today was first suggested formally by VICKERS, NETTLEBECK & WILLSON in 1972. The first study of the relationship between individuals' ITs and IQ scores was carried out in 1976 by Nettlebeck and Lally. They reported correlations of  $-0.89$  and  $-0.92$  between WAIS performance IQ and two estimates of inspection time. The study was small (10 subjects) and the IQ range was wide (from 47 to 119). Subsequent replications confirmed the finding (LALLY & NETTLEBECK, 1977; BRAND, 1981), and by 1982 BRAND & DEARY reviewed the nine known studies of IT and IQ and found a median correlation of  $-0.8$  in five studies where young adults of mean IQ around 100 had been tested with «culture-fair» tests. IT is an estimate of a person's speed of intake of sensory information. Practically, the task, as originally conceived, involved subjects viewing two matte black vertical lines of markedly different lengths in a tachistoscope. The lines, after a brief exposure, were backward masked by thicker matte black lines to prevent further examination of the stimulus for information. The task of the subject was to state whether the long line was the right or the left hand member of the pair of lines. The measure taken (the subject's IT) is the exposure duration at which the subject is able to perform the discrimination to a pre-set level of accuracy (often around 90%). No reaction time is taken in the standard task and subjects are encouraged to respond at leisure. Significant correlations have been found, in the expected direction, in IT tasks involving 2, 3 and 4 lines, 2 lights, and animal names as

stimuli. The correlation has been found in diverse subject groups: young children (BRAND & DEARY, 1982; ANDERSON, 1986; DEARY, 1987a); mentally handicapped and normal adult populations (BRAND & DEARY, 1982; NETTLEBECK, 1987); and in undergraduate samples where ability range is very restricted (MCKENZIE & BINGHAM, 1985; LONGSTRETH, WALSH, ALCORN, SZESZULSKI & MANIS, 1986; DEARY, 1987b).

The basic nature of the IT phenomenon is supported by the fact that the correlation holds for auditory as well as visual processing speed. In the auditory IT task the subject has to identify the temporal order of two tones («low-high» or «high-low») of markedly different pitch (usually separated by about 100 Hz). The tones are played for a brief time, one after the other, and are backward masked either by white noise or by a warble that contains the frequencies of both of the stimulus tones (DEARY, 1980, reported in BRAND & DEARY, 1982; NETTLEBECK, EDWARDS & VREUGDENHIL, 1986). The correlations between auditory and visual IT and IQ are similar: in a recent review NETTLEBECK found that of 24 independent studies (including auditory and visual stimuli), 16 found a significant negative correlation between IT and IQ. The correlations ranged from +0.1 to -0.61 with a mean of -0.34. Many of the studies involved groups that were of restricted ability range (and all the studies involving mentally retarded subjects were excluded), and Nettlebeck estimated a true IT-IQ correlation of around -0.5 in non-retarded young adults.

In a short time the IT measure has attracted much interest. Not all of this interest has been helpful in advancing the psychological understanding of the measure. The early hopes that IT might provide a widely-acceptable culture-fair estimate of ability were clearly not realised (BRAND & DEARY, 1982; NETTLEBECK, 1982; MACKINTOSH, 1981). However, the combination of the reliability and size of the IT-IQ relationship, the number of independent groups who have replicated it and the apparent simplicity of the task puts IT in a unique position for the study of basic processes in intelligence. The following points are a series of problems and possibilities that the measure has thrown up in its short existence.

### Problems and Prospects in Inspection Time Research

Individuals' strategies in performing the IT task have been examined (MACKENZIE & BINGHAM, 1985; MACKENZIE & CUMMING, 1986; EGAN, 1986; FITZMAURICE & NOLAN, 1983). When the two-lines task is performed using computer-drawn lines, or when lines are presented as a series of light-emitting diodes, as opposed to the tachistoscope lines, then about half of an undergraduate sample will report an apparent-movement artefact between the stimulus lines and the backward mask. They report using this impression of movement to solve the items. MACKENZIE and his colleagues have found that the IT-IQ correlation holds for the non-strategy-users but disappears in the strategy using group. Unpublished work in Edinburgh (EGAN, personal communication) has found both the strategy-users and the non users to have a similar IT-IQ correlation in a study involving a group with normal mean and standard deviation for IQ. However, the strategy involved here seems to be an artefact of computer- presentation devices and has not been reported in the tachistoscopic visual or with the auditory form of the test.

A more general approach to the strategy argument is put forward by those who claim that the high IQ person has a better IT by virtue of his willingness to sustain attention in a boring task. A reply to this criticism may be found in the observation that individuals' results, no matter how slow the IT, show regular psychometric curves. In other words, in the IT test, where exposure durations are varied unpredictably and are then, at the end,



plotted as a duration versus % correct, one would be hard put, using an attentional hypothesis, to explain how an «unattending» testee managed to score 100% at a duration of, say, 80ms and then fell off regularly until he or she was responding at chance levels when the duration came to 65ms. The subjects' hypothesised fatigue or unwillingness to respond because of boredom does not in fact result in the expected random error pattern. There is another objection to the attentional hypothesis. In a commonly used psychophysical IT procedure, the PEST (TAYLOR & CREELMAN, 1967) adaptive staircase, the subject has a short run of trials at one duration that suddenly jumps to a faster duration. If the IT task is simply a matter of attention then the items to be discriminated at these sudden duration jumps (when the duration gets shorter) would be less well solved by the low IT (poor attention) subjects. A re-analysis of data in this department by myself and Egan has shown that low IT and low IQ subjects cope with unexpected shortenings of duration just as well as those with shorter IT and higher IQ.

The cross-modal study of IT has added weight to its purported basic nature. Also, the auditory-visual correlation has been studied to find if the two tasks have some common variance, for it is possible that their separate correlations with IQ scores are a result of independent strategies. The original claim of a correlation of near to unity between auditory and visual inspection time (BRAND & DEARY, 1982) was dependent upon the inclusion of mentally retarded subjects and NETTLEBECK, EDWARDS & VREUGDENHIL (1986) and DEARY (1987b) subsequently found a correlation of about 0.4 for university samples. Given the restricted ability range of this population this suggests that there is a considerable amount of shared variance between the auditory and visual processing speeds and that this might reflect some property of the CNS. However, the correlation of IQ and processing speed of tactile information (EDWARDS, 1984) suggests that the result will not generalise to this third sense. But, given the cognitive unimportance of tactile as opposed to visual and auditory information perhaps that is not surprising.

The initial strength of the IT-IQ finding lay in the fact that it could be replicated on different hardware using different stimuli (this, of course, is not a weakness of the phenomenon as MACKINTOSH (1986) suggests). In his comprehensive review NETTLEBECK (1987) called for a concerted international effort to standardise the test. At present the IT is performed on tachistoscopes, computer screens and LEDs; experimenters have used lights, lines and sounds of different intensities and of different discriminabilities; and different psychophysical techniques have been used in its estimation (adaptive staircases and MCS methods as well as combinations of the two). In order that a subject's IT becomes more than an arbitrary number it will be necessary to standardise these many factors.

Meanwhile IT presents many advantages as a test of basic processing speed. It shows little practice effect (a reduction of 17-30% is typical over the first few sessions with very little improvement thereafter (NETTLEBECK, 1987) and may be used in repeated testing in pharmacological studies, time of day work, occupational testing and in longitudinal studies. It continues to remain one of the few correlates of general mental ability that has not been accused of some kind of bias. Unlike RT, IT has the appearance of a task that indexes one basic process: the speed of intake of sensory information. Unlike AEP, IT may be tested relatively quickly by workers using non-specialist equipment.

Moreover, IT offers the tantalising possibility of testing a human ability in animals such as rats, doves and non-human primates that are commonly used in the psychological laboratory. Indeed, perhaps that is the most important way forward for the IT measure in the biology of intelligence. Speculatively, it is possible to name three factors that might contribute to inter- and intra-species differences in what we call intelligence. Speed of

information processing is an obvious candidate: faster intake speed (as indexed by IT) will lead to enhanced ability to make sense of the incoming flow of information and will allow more detailed study of time-limited information sources (like road traffic or the motion of prey). Second, the degree of corticalisation will allow more complex and higher level thinking: allowing an animal to plan ahead with more foresight, to make more use of stored information and to organise incoming information in more detailed indexes. Third, there might be a place for the ability to make fine discriminations with the senses (DEARY, 1987a): a sensory system that has a smaller JND increment will accrue more information from the environment and will make fewer stimulus confusions. Here, then, is a proposed triad for the study of the biology of intelligence, both intraspecies and interspecies: sensory sensitivity, information processing speed and relative size of cerebral cortex.

## References

- ABELSON A.R., 1911. *The measurement of mental ability of backward children*. British Journal of Psychology, 4: 268-314.
- ANDERSON M., 1986. *Inspection Time and IQ in young children*. Personality and Individual Differences, 7: 677-686.
- BECK L.F., 1933. *The role of speed in intelligence*. Psychological Bulletin, 30: 169-178.
- BLINKHORN S.F. & HENDRICKSON D.E., 1982. *Average evoked responses and psychometric intelligence*. Nature, 295: 596-597.
- BORING E.G., 1923. *Intelligence as the tests test it*. New Republic, 35: 35-37.
- BRAND C.R., 1981. *General intelligence and mental speed: their relationship and development*. In: M.P. Friedman, J.P. Das and N. O'Connor, eds. Intelligence and Learning. Plenum.
- BRAND C.R. & DEARY I.J., 1982. *Intelligence and «Inspection Time»*. In: H.J. Eysenck, ed. A Model for Intelligence. Springer Verlag.
- BURT C., 1909-10. *Experimental tests of general intelligence*. British Journal of Psychology, 3: 94-177.
- CAMPBELL D.T., 1974. *Evolutionary epistemology*. In: P.A. Schlipp, ed. The Philosophy of Karl Popper, vol 1, pp. 413-436. Open Court, LaSalle.
- CANAVAN A.G.M., DUNN G. & McMILLAN T.M., 1986. *Principal components of the WAIS-R*. British Journal of Clinical Psychology, 25: 81-85.
- CAREY N., 1914-15. *Factors in the mental processes of schoolchildren*. British Journal of Psychology, 7: 453-490.
- CARYL P.G. & FRASER I.C., 1985. *The Hendrickson «string-length» measure and intelligence — a replication*. Paper presented at the Psychophysiology Society — 2nd Scottish meeting (September 1985).
- CATTALL J. McK, 1886. *The time taken up by cerebral operations*. Mind, 11: 220-242, 377-392, 524-538.
- CATTALL R.B., 1963. *Theory of fluid and crystallised intelligence: A critical experiment*. Journal of Educational Psychology, 54: 1-22.
- CHASE T.N., FEDIO P., FOSTER N.L., BROOKS R., DICHIRO G. & MANSI L., 1984. *WAIS performance: Cortical localisation by fluorodeoxyglucose F18-positron emission tomography*. Archives of Neurology, 41: 1244-1247.
- DEARY I.J., 1980. *How general is the mental speed factor in «general» intelligence*. B.Sc. Hons thesis, University of Edinburgh.
- DEARY I.J., 1986. *Inspection time: Discovery or rediscovery?* Personality and Individual Differences, 7: 625-631.
- DEARY I.J. 1987a. *The nature of intelligence: Simplicity to complexity and back again*. In: D. Forshaw and M. Shepherd, eds. The Maudsley Essay Series in the History of Psychiatry, vol 1. (In press).
- DEARY I.J., 1987b. *Visual and auditory inspection time: their interrelationship and correlation with IQ in high ability subjects*. Paper presented at the third biennial meeting of the International Society for the Study of Individual Differences, Toronto, June 1987.
- DEARY I.J., 1987c. *Intelligence, auditory inspection time and musical ability in schoolchildren*. Paper presented at the annual general meeting of the British Psychological Society, Scottish Branch, Rothesay, September 1987.

- DEARY I.J., HENDRICKSON A.E. & BURNS A., 1987. *Calcium and Alzheimer's disease: A finding and an aetiological hypothesis*. Personality and Individual Differences, 8: 75-80.
- EDELMAN G.M., & REEKE G.N. JR., 1982. *Selective networks capable of representative transformations, limited generalisations and associative memory*. Proceedings of the National Academy of Sciences U.S.A., 79: 2091-2095.
- EDWARDS C., 1984. *Inspection time in three sensory modalities and its relation to intelligence*. B.A. (Hons) Thesis, University of Adelaide.
- EGAN V., 1986. *Intelligence and inspection time: Do high-IQ subjects use cognitive strategies?* Personality and Individual Differences, 7: 695-700.
- ERTL J. & SCHAFER E.W.P., 1969. *Brain response correlates of psychometric intelligence*. Nature, 223: 421-422.
- EYSENCK H.J., 1973. *The Inequality of Man*. Temple Smith.
- EYSENCK H.J., 1979. *The Structure and Measurement of Intelligence*. Springer Verlag.
- EYSENCK H.J., (ed.), 1982. *A Model for Intelligence*. Springer Verlag.
- FITZMAURICE G. & NOLAN J.M., 1983. *Inspection time: A misguided model for intelligence*. Paper read to the Irish Psychological Society.
- FOLSTEIN M.F., FOLSTEIN S.E. & MCHUGH P.R., 1975. «Mini Mental State». *A practical method for grading the cognitive state of patients for the clinician*. Journal of Psychiatric Research, 12: 189-198.
- FREARSON W. & EYSENCK H.J., 1986. *Intelligence, reaction time (RT) and a new «odd-man-out» paradigm*. Personality and Individual Differences, 6: 807-818.
- GALTON F., 1883. *Inquiries into Human Faculty*. Dent.
- GOULD S.J., 1981. *The Mismeasure of Man*. Norton.
- GUILFORD J.P., 1985. *The structure-of-intellect model*. In: B.B. Wolman, ed. *Handbook of Intelligence*, pp 225-266. John Wiley Ltd.
- GUSTAFSSON J.E., 1984. *A unifying model for the structure of mental abilities*. Intelligence, 8: 179-203.
- HAIER R.J., ROBINSON D.L., BRADEN W. & WILLIAMS D., 1983. *Electrical potentials of the cerebral cortex and psychometric intelligence*. Personality and Individual Differences, 4: 591-599.
- HEARNshaw L., 1979. *Cyril Burt: Psychologist*. Cornell University Press.
- HENDRICKSON A.E. & HENDRICKSON D.E., 1982. *The biological basis of intelligence: Parts 1 and 2*. In: H.J. Eysenck, ed. *A Model for Intelligence*. Springer Verlag.
- HENDRICKSON D.E. & HENDRICKSON A.E., 1980. *The biological basis of individual differences in intelligence*. Personality and Individual Differences, 1: 1-33.
- HICK W., 1952. *On the rate of gain of information*. Quarterly Journal of Experimental Psychology, 4: 11-26.
- HOOPER F.H., HOOPER J.O. & COLBERT K.K., 1984. *Personality and memory correlates of intellectual functioning: Young adulthood to old age*. Karger.
- INOUE E., PARK K.S. & ASAKA A., 1984. *Blood uric acid level and IQ: A study in twin families*. Acta Genet. Med. Gemellol., 33: 237-242.
- JENSEN A.R., 1969. *How much can we boost IQ and scholastic achievement?* Harvard Educational Review, 39: 1-123.
- JENSEN A.R., 1980. *Bias in Mental Testing*. Methuen.
- JENSEN A.R., 1985a. *The nature of the black-white difference on various psychometric tests: Spearman's hypothesis*. Behavioural and Brain Sciences, 8: 193-263.
- JENSEN A.R., 1985b. *The plasticity of «intelligence» at different levels of analysis*. In: J. Lockhead, J. Bishop and D. Perkins, eds. *Thinking: Progress in Research and Teaching*. Franklin Institute Press.
- JENSEN A.R. & VERNON P.A., 1986. *Jensen's reaction-time studies: A reply to Longstreth*. Intelligence, 10: 153-179.
- KAMIN L.J., 1974. *The Science and Politics of IQ*. Lawrence Erlbaum Associates.
- LALLY M. & NETTLEBECK T., 1977. *Intelligence, reaction time and inspection time*. American Journal of Mental deficiency, 82: 273-281.
- LONGSTRETH L.E., 1986. *The real and the unreal: A reply to Jensen and Vernon*. Intelligence, 10: 181-191.
- MACKENZIE B. & BINGHAM E., 1985. *IQ, inspection time, and response strategies in a university population*. Australian Journal of Psychology, 37: 257-268.
- MACKENZIE B. & CUMMING S., 1986. *Inspection time and apparent motion*. Personality and Individual Differences, 7: 721-729.
- MACKINTOSH N.J., 1981. *A new measure of intelligence?* Nature, 289: 529-530.
- MACKINTOSH N.J., 1986. *The biology of intelligence?* British Journal of Psychology, 77: 1-18.



- NETTLEBECK T., 1982. *Inspection time: An index for intelligence?* *Quarterly Journal of Experimental Psychology*, 34A: 299-312.
- NETTLEBECK T., 1987. *Inspection time and intelligence*. In P.A. Vernon, ed. *Speed of Information Processing and Intelligence*. Ablex.
- NETTLEBECK T., EDWARDS C. & VREUGDENHIL A., 1986. *Inspection time and IQ: Evidence for a mental speed-ability association*. *Personality and Individual Differences*, 7: 633-641.
- NETTLEBECK T. & LALLY M., 1976. *Inspection time and measured intelligence*. *British Journal of Psychology*, 67: 17-22.
- PIAGET J., 1971. *Biology and Knowledge*. Edinburgh University Press.
- PIAGET J., 1978. *Behaviour and Evolution*. Routledge, Kegan and Paul.
- PIAGET J., 1980. *Adaptation and Intelligence*. University of Chicago Press.
- ROSE S., LEWONTIN R.C. & KAMIN L.J., 1984. *Not in Our Genes*. Pelican.
- SCHUCARD D.W. & HORN J.L., 1972. *Evoked cortical potentials and measurement of human abilities*. *Journal of Comparative and Physiological Psychology*, 78: 59-68.
- SINET P.M., LEJEUNE J. & JEROME H., 1979. *Trisomy 21 (Down's syndrome), glutathione peroxidase, hexose monophosphate shunt and IQ*. *Life Sciences*, 24: 29-33.
- SPEARMAN C., 1904. *«General intelligence», objectively determined and measured*. *American Journal of Psychology*, 15: 201-293.
- STERNBERG R.J., CONWAY B.E., KETRON J.L. & BERNSTEIN M., 1981. *People's conceptions of intelligence*. *Journal of Personality and Social Psychology*, 41: 37-55.
- STERNBERG R.J., 1985. *Human intelligence: The model is the message*. *Science*, 230: 1111-1118.
- TAYLOR M.M. & CREELMAN C.D., 1967. *Efficient estimate on probability functions*. *Journal of the Acoustical Society of America*, 4: 782-787.
- THURSTONE L.L., 1938. *Primary mental abilities*. *Psychometric Monographs*, No 1. University of Chicago Press.
- VICKERS D., NETTLEBECK T. & WILLSON R.J., 1972. *Perceptual indices of performance: The measurement of «inspection time» and «noise» in the visual system*. *Perception*, 1: 263-295.
- WEISS V., 1984. *Psychometric intelligence correlates with interindividual different rates of lipid peroxidation*. *Biomedica and Biophysica Acta*, 43: 755-763.
- WEISS V., 1986. *From memory span and mental speed toward the quantum mechanics of intelligence*. *Personality and Individual Differences*, 7: 737-749.
- WUKETITS F.M.P., 1986. *Evolution as a cognition process: Towards an evolutionary epistemology*. *Biology and Philosophy*, 1: 191-206.

Received: May 10, 1988. Accepted: May 28, 1988.

## VISUAL AND AUDITORY INSPECTION TIME: THEIR INTERRELATIONSHIP AND CORRELATIONS WITH IQ IN HIGH ABILITY SUBJECTS

IAN J. DEARY\*, PETER G. CARYL, VINCENT EGAN and DAVID WIGHT

Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, Scotland

(Received 5 May 1988)

**Summary**—Results from three visual inspection time (IT) tests and a new auditory IT test were correlated with psychometric measures of intelligence in an undergraduate population. The four IT tests correlated with each other at similar levels to the psychometric test intercorrelations. The range of raw IQ-IT correlations agreed with the predicted  $-0.35$  that has been reported previously. Corrected correlations indicated that the true IQ-IT correlation is in the region of  $-0.55$ . This study corroborates the hypothesis that a substantial proportion of the IQ variance is located in individual differences in perceptual intake speed, both visual and auditory. The intercorrelation of the auditory and visual IT tests indicates that IT is, in part, due to differences in nervous system functioning in general. The results do not support a specific strategy theory of IT performance.

### INTRODUCTION

It is more than 10 years since the first report of a high correlation between inspection time (IT) and measures of psychometric intelligence (Nettelbeck and Lally, 1976). IT is an estimate of the duration (in msec) needed by a subject in order to make a discrimination to a predetermined level of accuracy. In practical terms it often involves a subject reporting which of two lines of markedly different lengths is longer when the two are presented briefly. In Nettelbeck's (1987) estimation:

The best available estimate of the strength of the uncorrected correlation between IT and general ability among normal young adults (14 years and older) is  $-0.35$ .

This figure was extracted after analysing 16 representative IT studies which included 529 IT measures taken from 439 subjects. Nettelbeck calculated that the true value of this correlation, corrected for restriction of ability range, is about  $-0.5$ .

Despite the optimism expressed by some workers (Brand and Deary, 1982; Garnett, 1985) that IT would provide a culture fair, practice-resistant, convenient measure of intelligence, the measure is still in the development stage and there are many issues that are only partly resolved.

Most importantly, there is little evidence or agreement about the relationship between IT as it is measured in the visual and auditory modalities. This has both theoretical and practical importance: variance shared by the auditory and visual forms of the test indicates that some proportion of IT performance is due to basic differences in nervous system functioning; and being able to test more than one modality allows for testing in the blind and deaf. Deary (1980; reported in Brand and Deary, 1982) found that IT in the auditory modality correlated  $-0.66$  with verbal intelligence and  $-0.70$  with Raven's Matrices. Subjects, in the auditory test, were required to state the temporal order ('high-low' or 'low-high') of two tones. The tones were of 880 and 770 Hz, presented monaurally, separated by 500 msec of silence and were backward and forward masked with white noise. The reported correlation between visual and auditory IT was 0.99 but this correlation was dependent upon the inclusion of retarded subjects. Nettelbeck, Edwards and Vreugdenhil (1986) reported a correlation of  $-0.38$  between auditory IT and the Advanced Progressive Matrices in a university population. In the same study the correlation between visual and auditory IT was 0.39. Irwin (1984) found a correlation of 0.17 (Kendall's tau; Pearson's  $r = 0.05$ ) between auditory and visual IT tasks in 50 twelve year old children. Irwin has criticised the auditory measure and cautions that, as the duration of auditory stimuli decrease, the task

\*To whom correspondence should be addressed.

becomes more to do with pitch discrimination than processing speed. The correlation between auditory IT and pitch discrimination in Irwin's study was  $-0.51$  (Kendall's tau; Pearson's  $r = -0.54$ ). Irwin's study, however, contains no indication that there were subjects who could not perform the auditory IT task to a criterion level. In both Deary (1980) and Nettelbeck's (1987) experiments a significant minority of adults, including those of above average intelligence, were unable to make a reliable pitch discrimination even at very long durations. This emphasises the necessity for discrimination criteria and subjects who are not able to make a reliable discrimination should not be included in the analysis. If Irwin (1984) included all subjects without eliminating those who were "tone deaf" then the IQ-auditory IT and visual-auditory IT correlations would have been lowered.

Experimentally, Irwin's objections are answerable in two ways. First, the use of more effective masking for the auditory tones (Nettelbeck *et al.*, 1986) increases the stimulus durations at IT criterion levels and avoids the problem of frequency spectra overlapping when very short tones are presented. Second, the correlation between pitch discrimination, auditory IT and IQ tests can be performed independently. Deary (1980) administered a 42 item pitch discrimination test to his subjects. The correlation between pitch discrimination ability and Raven's Progressive Matrices scores was  $0.65$  ( $P < 0.05$ ); the Mill Hill correlation was  $0.45$  (NS). This does not include the retarded subjects. The correlation between pitch discrimination and auditory inspection time was  $-0.13$  (NS). This suggests that auditory processing speed and pitch discrimination correlate independently with intelligence. In attempt to resolve this issue, which we see as central to the mental speed theory of intelligence, we have devised an auditory IT task which overcomes some of the above difficulties. To prevent the use of echoic memory or rehearsal we have placed the two stimulus tones together temporarily; and for more effective masking we have used a warble composed of alternating 10 msec bursts of the two stimulus tones.

The aims of the present study were: (a) to develop a new auditory inspection time method which takes account of previous criticisms; (b) to establish the degree of the visual and auditory IT test intercorrelations; (c) to incorporate the different visual stimuli used in previous IT studies (Brand and Deary, 1982; MacKenzie and Bingham, 1985; and Longstreth, Walsh, Alcorn, Szeszulski and Manis, 1986) in a single study; and (d) to run IT tests as group tests.

## MATERIALS AND METHODS

### *Subjects*

Subjects were 120 second year psychology students taking part in a laboratory class. There were 81 women and 38 men. No sex differences have been found in IT studies and sex was not examined in our study. The mean age of the subjects was 20.39 years (SD 2.93). Subjects' first language was English and only those with adequate corrected or uncorrected vision and satisfactory hearing were used in the IT analyses.

### *Psychometric tests*

*Advanced Progressive Matrices, 1962 revision* (Raven, Court and Raven, 1977). This was chosen as a non-verbal test for the discrimination of high ability subjects. Both sets were used. Set I is not scored and has 12 items that serve to familiarise the subjects with the form of the items. Set II was used as a timed (40 minute) test. It has 36 items of increasing difficulty.

*Mill Hill Vocabulary Form 1 Senior—Synonyms* (Raven, Court and Raven, 1977). This provided a quick test of vocabulary level but is intended for a general population.

*Alice Heim 5* (Heim, 1968). This is a two part ability test which discriminates between high ability subjects. Part A is verbal-numerical. Part B is non-verbal. Each section has a time limit of 20 min.

### *Inspection time tests*

*Visual.* Three visual IT tests were devised for this study. Stimuli were presented on a micro-computer screen. Duration of each stimulus was decided by a programmed version of the PEST adaptive staircase method (Taylor and Creelman, 1967). At the end of the run this provided an IT estimate for a subject at the 85% level of discrimination accuracy. Easy items at the beginning of each IT test provided the only practice and served to familiarise the subject with the requirements

of the tests. In each of the visual tasks there was a rest after 10 trials. The subject initiated the restart. If the subject had performed at 80% or better in the last 5 items the message 'Well Done' appeared on the screen. If the subject had scored less than 80% correct in the last 5 items the message 'Pay Close Attention' appeared. All stimuli, screen backgrounds and general background light levels were checked for equality using a light meter. Monitor controls were made inaccessible to the subjects.

All three visual tests were designed to be self administered and self-paced. Each trial had a similar sequence. A cue (identical to the backward mask) lasting 500 msec preceded the stimulus by 1 sec. Immediately after the stimulus the backward mask lasted for 1 sec. Subjects responded at their leisure. A response initiated the next item.

*\*Vertical Lines Test.* The first visual IT test used the familiar vertical lines of markedly different lengths. The long line was 5 cm, the short line was 3 cm, they were 2 cm apart and they were approximately 2 mm thick. The mask was a pair of identical lines 7 cm long and 5 mm thick that entirely covered the stimulus lines. Subjects were required to indicate, by pressing one of two computer keys, whether the long line (stimulus presentations always included one long line and one short line) was on the left or the right. In this test, as with all other IT tests here, a record is kept of the correctness of each response and no record is kept of the response time since it is emphasised that subjects should respond slowly and with maximum accuracy.

*\*Horizontal Lines Test.* This was a modification of the test used by MacKenzie and Bingham (1985). The stimuli were two horizontal lines of the same lengths as the vertical lines. The masking lines were 7 cm long. The lines were about 1 mm thick, as were the masking lines. Unlike the MacKenzie and Bingham stimuli the lines were presented side by side. Subjects' tasks and responses were essentially the same as for the vertical lines test.

*\*'Longstreth' Task.* This test was similar to that used by Longstreth *et al.* (1986). The stimulus was either a diagonal slash about 1 cm long or a rectangle of 8 mm by 6 mm. The mask was a combination of the two stimuli that occupied exactly the same area of the screen. Subjects were asked to indicate, by pressing one of two keys, whether they had seen the diagonal slash or the rectangle prior to the mask. Note that this IT test, unlike the previous two, does not involve the comparison of two simultaneously presented stimuli.

*Auditory test.* The auditory test was a fixed-pace test using the method of constant stimulus durations. Stimuli were square wave tones produced using an XR320 monolithic timing circuit driven by a BBC microcomputer. Stimulus tones were 880 Hz (high) and 784 Hz (low). The auditory mask was a warble of both tones (10 msec each) provided by a multivibrator circuit. Test stimuli were presented using a UHER reel to reel tape recorder with 1/4 inch tape run at high speed. The output was relayed to 18 headphone sets in individual quiet basement cubicles. Sound levels were equalised for all headsets at about 80 dB for stimuli and mask. Each auditory trial consisted of a cue tone (832 Hz) lasting 500 msec, 1 msec of silence, a stimulus pair ('high-low' or 'low-high') of tones of given identical durations followed by the warble mask for 1 sec. There was no gap between the stimuli and the mask. About 8 sec of silence between items gave subjects adequate time to tick 'high-low' or 'low-high' on a response sheet. All instructions for this test were on the pre-recorded tape. Two practice blocks, 10 trials each, followed. These provided an introduction to the task at easy stimulus durations (200 msec). The experimental session consisted of thirteen blocks of 10 stimulus pairs presented at decreasing durations—200, 150, 125, 100, 85, 70, 55, 40, 30, 20, 15, 10 and 6 msec. Subjects were required to score 90% or above in the second practice and first experimental blocks before being considered able to discriminate pitch. (No subject who failed at 200 msec went on to achieve 90% or greater at more difficult durations and no subject had an auditory IT in the 150–200 msec range indicating that the subjects who failed this criterion were tone deaf for the discrimination required.)

### Procedure

Subjects were randomly allocated to two groups. Group 1 took the psychometric tests before the IT session. Group 2 completed the sessions in the reverse order. For each subject all estimations



were performed in a single afternoon session. (Except AH5 which had been completed by a random subset of 60 subjects in afternoon sessions 2 months previously.)

The psychometric tests were administered as group tests under examination conditions according to the instructions in the respective manuals.

For the IT session subjects were given a general briefing of the order of events, a general description of the visual tests and instructions on how to load and administer the tests for themselves. (All subjects were familiar with the microcomputer network from previous practical sessions.) Subjects, in groups of 18, went to individual cubicle rooms and put on a set of headphones. They listened to the pre-recorded auditory IT tape which gave full instructions for test completion. The cue, stimuli and mask were introduced separately and specimen items were built up sequentially. Several items were then played with the appropriate answers ("high-low" or "low-high"). Practice blocks 1 and 2 and the 13 experimental blocks followed. There was a short pause between each block. The answer to each item was entered in a response sheet. Total testing time for the auditory IT test was 40 min.

After the auditory test the subjects administered the visual IT tests to themselves. All subjects were instructed to perform the vertical lines test. Due to time constraints, subjects performed only one of the two other visual IT tests. Following the vertical lines test, after a short rest, half of the subjects performed the horizontal lines test and half of the subjects performed the Longstreth task.

### Scoring

On completion of each visual IT test the monitor screen presented the subject with printed feedback about his/her performance in the form of a number representing the duration at which the subject was 85% correct and the number of trials completed. Data for each trial was also collected on network files. The auditory test was scored by hand from the response sheets. The auditory IT was said to be the last duration at which the subject was 90% correct for that and all easier durations. Errors at longer stimulus durations could be recouped by 100% success at more difficult durations.

## RESULTS

All subjects provided Mill Hill Vocabulary (MHV) scores. Advanced Progressive Matrices (APM) scores were obtained from 119 subjects due to a spoiled answer sheet. Alice Heim 5 (AH5) scores were available for 60 subjects. After eliminating subjects with unsatisfactory vision and anonymous response sheets 105 Vertical Lines test results were available; 51 results were collected for the Horizontal Lines test and 50 did the Longstreth test. Auditory IT results were available for 119 subjects. Of these, 80 subjects met the criteria for IT estimation. In this high ability group 32.8% of subjects found the discrimination (two tones separated by 96 Hz) too difficult.

Mean and standard deviations for all measures are given in Table 1. The APM mean of 24.9 corresponds to an IQ mean of about 124 with a standard deviation of about 7 points. This is similar to, if slightly higher than, the previously reported means for undergraduates (Raven, Court and Raven, 1977). The MHV score of 31.9 (SD 3.46) is not directly comparable to population means as it is only a synonym test but it shows a similarly restricted range to the APM. The AH5 test discriminates this population more successfully than either of the other two tests and the mean of 36.4 (total score for Parts A and B) and standard deviation of 7.45 are close to those reported on

Table 1. Means and standard deviations for the tests of psychometric intelligence and inspection time

Test	<i>n</i>	Mean	Standard deviation
Advanced Progressive Matrices	119	24.9	3.46
Mill Hill Vocabulary	120	31.9	3.46
Alice Heim 5 (Test A)	60	16.9	3.78
Alice Heim 5 (Test B)	60	19.6	4.76
Alice Heim 5—total score	60	36.4	7.45
Vertical Lines IT	105	43.6 msec	27.9
Horizontal Lines IT	51	75.4 msec	29.2
Longstreth IT	50	29.8 msec	29.3
Auditory IT	80	75.8 msec	27.5

Table 2. Pearson's  $r$  correlations among different tests of psychometric intelligence

	AH5(A)	AH5(B)	AH5 (total)	Mill Hill vocabulary	Advanced progressive matrices
AH5(A)	—	0.50 ( $N = 60$ ) $P = 0.000$	0.83 ( $N = 60$ ) $P = 0.000$	0.31 ( $N = 60$ ) $P = 0.015$	0.42 ( $N = 59$ ) $P = 0.001$
AH5(B)		—	0.87 ( $N = 60$ ) $P = 0.000$	0.46 ( $N = 60$ ) $P = 0.000$	0.48 ( $N = 59$ ) $P = 0.000$
AH5(Total)			—	0.43 ( $N = 60$ ) $P = 0.001$	0.52 ( $N = 59$ ) $P = 0.000$
Mill Hill vocabulary				—	0.25 ( $N = 119$ ) $P = 0.007$

AH5 is the Alice Heim 5 test. Test A is verbal and numerical. Test B is non-verbal. Mill Hill vocabulary is the Form 1 Senior Synonyms test.

previous groups of university undergraduates (Heim, 1968). Thus the subject population is equivalent to other undergraduate samples and has a very low standard deviation. We should expect this to reduce all IT-IQ correlations.

All three visual tests gave similar standard deviations which indicates that they were discriminating among the subjects to a similar degree (Table 1). The Vertical Lines IT task was relatively easy (mean 43.6 msec, SD 27.9): 18 subjects (17.1%) scored at or better than the speed limit (20 msec) of the monitor screen. Twenty subjects (40%) managed this in the Longstreth task (mean 29.8 msec, SD 29.3). The Horizontal Lines IT task had the highest mean value of 75.4 msec (SD 29.2). Only 1 subject did better than 20 msec on the Horizontal Lines test.

In the auditory task the warble mask and the elimination of a gap between the auditory stimulus pairs has had the effect of elevating the mean auditory IT from the levels previously reported by us (Brand and Deary, 1982). The mean auditory IT was 75.8 msec (SD 27.5). The auditory task has a similar standard deviation to the three visual tasks.

Table 2 contains the Pearson  $r$  correlations among the psychometric tests. All correlations are in the expected direction and are significant. The MHV test provides the two lowest correlations: 0.25 with the APM and 0.31 with the verbal-numerical section of the AH5 test. Ignoring the within-AH5 correlations the others fall into the narrow range of 0.42 to 0.52.

Table 3 provides the correlations among the various tests of IT. All correlations are positive and all but one are significant. They fall into a similar range when they are compared with the inter-psychometric test correlations. The standard vertical lines task correlates at: 0.48 with the Horizontal Lines test; 0.39 with the Longstreth test; and 0.24 with the auditory test. The auditory test correlates at: 0.20 with the Horizontal Lines test; and at 0.53 with the Longstreth test. Note that despite the pile-up of maximum scorers in the Vertical Lines and Longstreth tests the correlations between Horizontal Lines and Longstreth with the classical Vertical Lines test are very significant and achieve levels in the upper reaches of the inter IQ test correlations.

Table 4 has the correlation coefficients between the four IT measures and the ability tests. Only one correlation is not in the expected direction and this, even when corrected, is near to zero. The test with the largest standard deviation and, by implication, the best discriminating power achieves the most consistent results. The AH5 total correlates with all four IT measures at very similar levels,

Table 3. Pearson's  $r$  correlations among the four tests of inspection time

	Vertical Lines	Horizontal Lines	Longstreth	Auditory
Vertical Lines	—	0.48 ( $N = 51$ ) $P = 0.000$	0.39 ( $N = 46$ ) $P = 0.007$	0.24 ( $N = 68$ ) $P = 0.045$
Horizontal Lines		—	Not tested	0.20 ( $N = 32$ ) $P = 0.27$
Longstreth			—	0.53 ( $N = 33$ ) $P = 0.001$



Table 4. Pearson's *r* correlations between the tests of psychometric intelligence and inspection time

	AH5(A)	AH5(B)	AH5 (Total)	Mill Hill vocabulary	Advanced progressive matrices
Vertical Lines	-0.31 ( <i>N</i> = 51) <i>P</i> = 0.025	-0.27 ( <i>N</i> = 50) <i>P</i> = 0.06	-0.33 ( <i>N</i> = 51) <i>P</i> = 0.020	-0.02 ( <i>N</i> = 50) <i>P</i> = 0.82	-0.16 ( <i>N</i> = 104) <i>P</i> = 0.11
Horizontal Lines	-0.22 ( <i>N</i> = 20) <i>P</i> = 0.36	-0.32 ( <i>N</i> = 20) <i>P</i> = 0.17	-0.32 ( <i>N</i> = 20) <i>P</i> = 0.17	-0.17 ( <i>N</i> = 51) <i>P</i> = 0.24	-0.23 ( <i>N</i> = 50) <i>P</i> = 0.11
Longstreth	-0.06 ( <i>N</i> = 59) <i>P</i> = 0.73	-0.41 ( <i>N</i> = 32) <i>P</i> = 0.021	-0.29 ( <i>N</i> = 32) <i>P</i> = 0.10	-0.31 ( <i>N</i> = 50) <i>P</i> = 0.031	-0.28 ( <i>N</i> = 59) <i>P</i> = 0.048
Auditory	-0.11 ( <i>N</i> = 40) <i>P</i> = 0.49	-0.40 ( <i>N</i> = 40) <i>P</i> = 0.010	-0.31 ( <i>N</i> = 40) <i>P</i> = 0.055	-0.27 ( <i>N</i> = 80) <i>P</i> = 0.016	0.05 ( <i>N</i> = 80) <i>P</i> = 0.65

AH5 is the Alice Heim 5 test. AH5(A) is verbal and numerical and AH5(B) is non-verbal. Mill Hill Vocabulary is the Form 1 Senior Synonyms test.

between  $-0.29$  and  $-0.32$ . Only two of these are significant (one is a trend) but the closeness of all four correlations suggests a true result for this population. The AH5 non-verbal subtest achieves a similar range of correlations and exceeds the AH5 total with the auditory and Longstreth tests ( $-0.40$  and  $-0.41$ , respectively). These two IT tests have the highest intercorrelation and are also the IT tests that correlate best with those psychometric tests that have lower discriminating power (APM and MHV)—see Table 4.

Taking the AH5 total, the correlation between IQ and IT is close to Nettelbeck's  $-0.35$ . We should expect the IT-IQ correlation to be reduced in this undergraduate population because of the low range of ability. After correction for the reduction of variance (McNemar, 1955) the correlation values cluster around  $-0.5$  (Table 5). Again, the AH5 test suggests an even higher correlation, between  $-0.55$  and  $-0.6$ .

When the results of the Longstreth, Vertical Lines and Auditory tasks are included in a multiple correlation with the Alice Heim 5 total score the multiple correlation is  $0.94$  ( $N = 20$ ,  $P = 0.0004$ , adjusted  $R^2 = 0.87$ ). When the same tests are included in a multiple correlation with the other high ability test, the Raven's APM, the multiple correlation is  $0.47$  ( $N = 26$ ,  $P = 0.009$ , adjusted  $R^2 = 0.19$ ).

## DISCUSSION

This study has shown that the IT-IQ relationship is amenable to analysis under the conditions (group testing, restricted variance in IQ and high mean IQ) in which the parameters of the tests will be most readily studied by experimental psychologists. It also establishes that the correlation between IQ and IT can be obtained despite the fact that in some tests many of these high IQ subjects were able to reach or exceed the maximum level of performance that the method of presentation could discriminate. The results challenge Mackintosh's (1986), Todman and Gibb's (1985) and Vernon's (1986) suggestion that the IT-IQ correlation depends on the inclusion of retarded subjects. When the IT-IQ correlations were corrected for restriction of ability range, the correlations became slightly higher than the  $-0.5$  result predicted by Nettelbeck (1987). Within our particular ability range, the IT correlations for all four IT tasks with psychometric tests were close to the  $-0.35$  level that Nettelbeck estimated for normal adults. With our large samples many low correlations were clearly significant. However, if we had chosen a small sample size typical of

Table 5. Correlations *r* between tests of psychometric intelligence and inspection time corrected for restriction of ability range

	AH5(A)	AH5(B)	AH5(C)	Mill Hill vocabulary	Advanced progressive matrices
Vertical Lines	-0.57	-0.52	-0.60	-0.04	-0.16
Horizontal Lines	-0.44	-0.59	-0.59	-0.36	-0.45
Longstreth	-0.13	-0.69	-0.54	-0.57	-0.53
Auditory	-0.22	-0.68	-0.57	-0.52	0.11

many IT studies (e.g.  $N = 20$ ), we can calculate that, as a consequence of small sample size and reduced variance, we would have failed to find correlations significant at the 5% level in approximately 70 out of every 100 experiments (see Table 6). The implication is clear: while the presence of a significant correlation corroborates the IT-IQ relationship, the absence of one may not be used as evidence against it unless the sample size is large and the variance not restricted.

The IT tests, for the first time, were run as group tests and the time saved on the normal IT testing procedure was considerable. The tests were self-administered and, beyond familiarising the subjects with the stimuli and the computer and screen, there was minimal practice involved. Again, these changes cut down on the time of testing and increased the convenience of the tests. Despite the lack of practice the IT tests correlated as well together as did the IQ tests where, typically, the test-retest reliability is in the region of 0.8 and above. Prolonged practice might be unnecessary in IT testing given that the subjects are equally naive.

The issues addressed by the test were not primarily methodological. The study adds to those few that have examined auditory and visual IT with intelligence in the same experiment. Our findings are clear. Auditory IT ability correlates with ability tests at similar levels to the visual tests. The auditory test correlates significantly with the visual form of the test. The present result must be added to the original report of a correlation of 0.99 between visual and auditory IT tests (Deary, 1980), Edwards' (1984) correlation of 0.39 (unlike Deary's study, Edwards included only non-retarded subjects) and the near zero results reported by Irwin (1984). At this stage the best estimate of the correlation between visual and auditory IT in an adult population is in the region of 0.4. Although this is probably depressed by restriction of range in subject samples it would be unwise to estimate a 'true' value until larger scale studies have examined populations with normal distributions of ability. The suggestion from this is that the two sensory intake speeds share some common variance whose seat might be in the biology of the systems. One candidate as a source of this shared variance in neural efficiency is the number of errors made in synaptic transmission (Hendrickson and Hendrickson, 1982). Others have suggested that the biological differences in nerve conduction speed and synaptic delay might provide a basis for individual differences in ability (Reed, 1984, 1986). These views provide plausible biological bases for the psychological speed indexed in the IT measures. Reed's formulation is simpler: that psychological speed derives from faster neurological transmission. The Hendrickson hypothesis is more subtle: psychological speed, here, derives from the integrity of information transfer across synapses and transmission along axons. Clinical studies tend to support the latter view. Psychological speed is often slowed in cases where there is widespread injury to brain tissue as in closed head injury (Gronwall and Wrightson, 1981), Korsakoff's psychosis and Alzheimer's syndrome (Kopelman, 1985).

With the range of ability tests and IT tests used in the present study we hoped to be able to investigate the differences in IT-IQ correlations in verbal and non-verbal tests. Nettelbeck (1987) found that no specific type of intelligence had higher correlations with IT than others. Including only those studies carried out using non-retarded adults he reported that: of 12 studies using verbal ability estimates, the average correlation with IT was  $-0.27$  (7 were significant in the expected direction); of 9 studies using performance ability estimates, the average correlation was  $-0.33$  (7 were significant in the expected direction); and of 24 studies examining 'general' intelligence, the average correlation was  $-0.34$  (16 were significant in the expected direction). One recent study (Cooper, Kline and Maclaurin-Jones, 1986), which investigated the relationships between visual IT and the primary abilities, indicated that IT was more related to factors of visualisation and perceptual speed than to crystallised or fluid ability. That study, however, was based on results from only 20 undergraduates and the authors' conclusions are impressionistic and not based upon a factor analysis. The results of the present study are in agreement with the findings of Nettelbeck's review. Both auditory and visual IT tests correlated significantly with both verbal and non-verbal ability scores. The auditory IT test correlations are especially interesting: AIT correlates significantly with the non-verbal section of the Alice Heim and the Mill Hill vocabulary scores yet it correlates at near zero levels with verbal section of the Alice Heim and the Raven's APM. When the Longstreth test correlations are examined (Table 4) there are significant correlations with Mill Hill, Raven and AH5 (non-verbal) scores, but a near zero correlation with AH5 (verbal). The most parsimonious explanation of these results is that there is some general factor being tapped by IT and that, given the size of some of our subject subpopulations, we have no more non-significant

Table 6. Numbers of significant correlations obtained with different sized samples given that two members have a known amount of shared variance

$r$	$r^2$	$r$ Mean simulated	% of samples failing to reach $P = 0.05$ level of significance				
			$N = 20$	$N = 30$	$N = 40$	$N = 50$	$N = 100$
0.224	0.05	0.21	81	89	75	65	43
0.316	0.10	0.30	66	62	53	45	11
0.387	0.15	0.38	64	42	28	18	4
0.447	0.20	0.45	51	30	16	9	1
0.500	0.25	0.51	39	8	7	1	0
0.584	0.30	0.54	24	10	1	0	
0.633	0.40	0.62	13	3	2		
0.707	0.50	0.70	6	0	0		

To obtain an approximate value for the proportion of experiments in which a correlation genuinely present in the population would fail to reach significance (at  $P < 0.05$ ) because of the use of a small sample size we devised a simulation method. For each value of sample size ( $N$ ) shown above (Table 6), we simulated 100 IQ and IT samples from a normal population with a proportion of shared variance ( $r^2$ ) set at the values shown (Caryl, in preparation). For each pair of simulated samples, we calculated the Pearson correlation coefficient, and tallied the number of these which reached or exceeded the critical value for significance (at  $P = 0.05$ ) for a sample of the given size.

correlations than might be expected (see Table 6). This will require a much larger study but we predict that there is no particular type of intelligence being tapped by IT: it measures a general mental speed whose operational qualities are indexed by fluid or performance measures and whose products are estimated by verbal or information tests.

The present study did not set out specifically to address the subject of strategy use in performing the IT tests. Certainly, the visual tests have been criticised for being vulnerable to strategies. One of the most cited is the apparent motion strategy whereby the mask overlaying the stimulus gives the impression of movement and this movement may be used by subjects to estimate the position of the long and short lines in the vertical lines task. MacKenzie and Bingham (1985) and Egan (1986) have discussed this form of strategy use which we call the specific strategy theory. This asserts that within a single task or within a single modality there might be a 'short cut' to the solution of the problem. Such specific strategy theories have problems in explaining our results. It would be difficult to envisage a specific strategy that was useful in all three of the visual tests and in the auditory version of the test. Our study, however, is not capable of answering the more general strategy theorist whose claim is that the high IQ person is more attentive, vigilant, motivated or generally more organised in his approach to the IT tests. Given the regularity of our results we feel that it falls to the general strategy theorist to formulate a testable hypothesis before further consideration is given to this possibility.

IT tests, though, continue to have their drawbacks. The PEST method can reduce testing time per session to between 12 and 20 min but results in subjects having variable numbers of trials per session. (We found no significant correlations between IT estimates and trial numbers.) The method of constant stimulus durations gives a more standard test but subjects find this boring and, in contrast with the adaptive staircase, the form of the test fails to give the impression of periodically 'easing off'. We found that the computer screens, especially in the Longstreth test, could not provide a fast enough presentation to estimate all ITs beyond the failure point. It is important to emphasise that, even with 40% of subjects receiving the same score on this test, the correlation still held.

A persistent problem exists with the auditory measure. Subjects find the pitch discrimination difficult. We have little hope that a test which cannot be done by almost a third of undergraduates will be helpful for the general population.

Nevertheless, for those subjects who could do the test their ability scores correlated at moderate levels with auditory inspection time. It appears that the inability to do this test is independent of general hearing and intellectual ability and it might be time to try an alternative to pitch discrimination in IT testing. There is a large literature on auditory backward masking and alternatives such as loudness, duration or timbre might provide better alternatives (Kallman, Hirtle and Davidson, 1986).

Choice of IT tests may turn out to be important, or a variety of alternative tests may provide equally useful measures. This can only be established once the role of the various stimulus

parameters has been fully investigated. But the need for analysis of the extent to which choice of parameter can affect the 'fine tuning' should not detract from the consistency of the IQ-IT relationship observed here across a variety of IT tests.

*Acknowledgement*—The authors would like to thank Mr C. R. Brand for providing Alice Heim data for some subjects.

## REFERENCES

- Brand C. R. and Deary I. J. (1982) Intelligence and "inspection time". In *A Model for Intelligence* (Edited by Eysenck H. J.). Springer, Berlin.
- Cooper C., Kline P. and MacLaurin-Jones L. (1986) Inspection time and primary abilities. *Br. J. Educ. Psychol.* **56**, 304-308.
- Deary I. J. (1980) How general is the mental speed factor in 'general' intelligence: An attempt to extend inspection time to the auditory modality. B.Sc. (Hons) thesis, Univ. of Edinburgh.
- Edwards C. (1984) Inspection time in three sensory modalities and its relation to intelligence. B.A.(Hons) thesis, Univ. of Adelaide.
- Egan V. (1986) Intelligence and inspection time: Do high IQ subjects use cognitive strategies? *Person. individ. Diff.* **7**, 695-700.
- Garnett P. D. (1985) Intelligence measurement by computerised information processing using inspection time. *J. gen. Psychol.* **112**, 325-335.
- Gronwall D. and Wrightson P. (1981) Memory and information processing capacity after head injury. *J. Neurol. Neurosurg. Psychiat.* **44**, 889-895.
- Heim A. (1968) *Manual for the AHS Group Test of High Grade Intelligence*, Revised edn. NFER-Nelson, Windsor.
- Hendrickson A. E. and Hendrickson D. E. (1982) The biological basis of intelligence. In *A Model for Intelligence* (Edited by Eysenck H. J.). Springer, Berlin.
- Irwin R. J. (1984) Inspection time and its relation to intelligence. *Intelligence* **8**, 47-65.
- Kallman H. J., Hirtle S. C. and Davidson D. (1986) Recognition masking of auditory duration. *Percept. Psychophys.* **40**, 45-52.
- Kopelman M. D. (1985) Rates of forgetting in Alzheimer-type dementia and Korsakoff's syndrome. *Neuropsychologia* **23**, 623-638.
- Longstreth L. E., Walsh D. A., Alcorn M. B., Szeszulski P. A. and Manis F. R. (1986) Backward masking, IQ, S.A.T. and reaction time: Interrelationships and theory. *Person. individ. Diff.* **7**, 643-651.
- MacKenzie B. and Bingham E. (1985) IQ, inspection time and response strategies in a university population. *Aust. J. Psychol.* **37**, 257-268.
- Mackintosh N. J. (1986) The biology of intelligence? *Br. J. Psychol.* **77**, 1-18.
- McMemar Q. (1955) *Psychological Statistics*, 2nd edn. Wiley, New York.
- Nettelbeck T. (1987) Inspection time and intelligence. In *Speed of Information Processing and Intelligence* (Edited by Vernon P. A.). Ablex, Norwood, N.J.
- Nettelbeck T., Edwards C. and Vreugdenhil A. (1986) Inspection time and IQ: Evidence for a mental speed-ability association. *Person. individ. Diff.* **7**, 633-641.
- Nettelbeck T. and Lally M. (1976) Inspection time and measured intelligence. *Br. J. Psychol.* **67**, 17-22.
- Raven J. C., Court J. H. and Raven J. (1977) *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Lewis, London.
- Reed T. E. (1984) Residual latency (delay at the neuromuscular junction): Normative values and heritability in mice. *Behav. Genet.* **14**, 209-219.
- Reed T. E. (1986) A neurophysiological basis for the heritability of vertebrate intelligence. Paper presented at the NATO Advanced Study Institute on the Evolutionary Biology of Intelligence, Poppi, Italy, July.
- Taylor M. M. and Creelman C. D. (1967) PEST: Efficient estimates in probability functions. *J. acoust. Soc. Am.* **4**, 782-787.
- Todman J. and Gibb C. McD. (1985) High speed memory scanning in retarded and non-retarded adolescents. *Br. J. Psychol.* **76**, 49-57.
- Vernon P. A. (1986) Inspection time: Does it measure intelligence? *Person. individ. Diff.* **7**, 715-720.



# Auditory Inspection Time, Intelligence and Pitch Discrimination

IAN J. DEARY

*University of Edinburgh, Scotland*

BRIAN HEAD

*The Mary Erskine School and Daniel Stewart's and Melville College,  
Edinburgh, Scotland*

VINCENT EGAN

*University of Edinburgh, Scotland*

Irwin (1984) suggested that the correlation between auditory inspection time (AIT) and IQ was due to AIT being related to pitch discrimination ability. Previous forms of the AIT test have confounded temporal resolution speed and pitch discrimination. In this study, 59 undergraduates and 119 schoolchildren were tested on a new AIT test, pitch discrimination tests and verbal and nonverbal mental ability tests. AIT and IQ correlated at  $-.45$  (verbal) and  $-.27$  (nonverbal) in undergraduates and at  $-.36$  (verbal) and  $-.26$  (nonverbal) in children. There was a small but significant correlation between pitch discrimination and IQ scores in children but not in undergraduates. Pitch discrimination tended to correlate with AIT in children but not in undergraduates. Children with high verbal ability appear to have superior auditory inspection times and pitch discrimination abilities. When the effects of pitch discrimination were partialled out, the AIT-IQ correlations were altered very little in either sample. We conclude that the AIT-IQ correlation is due to the AIT being an index of information intake speed. AIT appears not to develop (i.e., to decrease) from age 12.5 to 21.

## INTRODUCTION

The hypothesis that individual differences in cognitive ability, as indexed by IQ-type tests, are related to differences in speed of information processing has received much recent corroboration. There are three main indices of processing speed (Mackintosh, 1986) which are found to correlate with ability test scores; namely, various measures of reaction time (e.g., Jensen & Vernon, 1986), components of brain average evoked potentials (Blinkhorn & Hendrickson, 1982), and inspection time (IT) (Nettelbeck, 1987; Vickers, Nettelbeck, & Willson, 1972). IT is a measure of the minimum stimulus duration that a subject requires in order to make a decision about a stimulus to a predetermined level of

---

Correspondence and requests for reprints should be sent to Ian J. Deary, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, Scotland.

accuracy. Following Vickers' (1970) original design the discrimination most commonly involves a subject judging which of two vertical lines of markedly different lengths is longer. The stimulus is immediately backward masked after presentation and the subject responds at leisure. The average correlation between intelligence test scores and IT measures over more than 20 studies is about  $-.35$  (Nettelbeck, 1987). When a correction is made for the restriction of ability range that is common to many samples (due to the number of samples that are composed of undergraduates) the true IT-IQ correlation is approximately  $-.05$  (Nettelbeck, 1987).

Questions have been raised about the generality of the IT-IQ correlation. Specifically, there have been suggestions that visual IT may involve the detection of a movement artifact in the stimulus (Mackenzie & Bingham, 1985; Mackenzie & Cumming, 1986). To discover whether the correlation of information intake speed (as indexed by IT) with IQ was a general phenomenon, not limited to one task or one modality, an auditory IT (AIT) task was devised (Deary, 1980, reported in Brand & Deary, 1982). AIT involved subjects discriminating the temporal order (high-low or low-high) of two briefly-presented square wave tones of markedly different pitch (770 and 880 Hz). The tones were forward and backward masked using white noise and were separated by a gap of 500 ms. AIT, the tone presentation time required by subjects in order to make a 90% accurate or better judgment about temporal order, correlated at  $-.70$  with Raven's Progressive Matrices scores and  $-.66$  with Mill Hill Vocabulary scores in a small group of subjects ( $n=13$ ), two of whom were mentally handicapped.

There have been few studies of AIT to date, but it is clear that the original AIT task could be improved. The auditory inspection times reported by Brand and Deary (1982) ranged from 6 to 160 ms, but the vast majority of the scores fell between 6 and 20 ms (median 10 ms). Irwin (1984) used an AIT task similar to that of Deary (1980) to test 50 children (mean age 12 years 2 months) and reported a mean AIT of 195.6 ms with a standard deviation of 339.7 and a median of 16 ms. Irwin reported IQ-AIT correlations (Kendall's tau) of  $-.23$  ( $p<.05$ ) for Raven's Progressive Matrices and  $-.32$  ( $p<.01$ ) for Mill Hill Vocabulary scores. Nettelbeck, Edwards, and Vreugdenhil (1986) devised an improved AIT task. They used 880 and 770 Hz tones backward and forward masked by a noise which consisted of alternating 15 ms bursts of the two stimulus tones and white noise. The masker noise sounded for 1000 ms before, between, and following the stimulus tones. Stimuli were presented using the PEST algorithm (Taylor & Creelman, 1967). Thirty subjects were tested on AIT and Raven's Advanced Progressive Matrices (IQ range 102-135,  $M$  121,  $SD$  8) and the AIT-IQ correlation was  $-.38$  ( $p<.05$ , one-tailed). In the same experiment auditory and visual IT correlated at  $0.39$  ( $p<.05$ ), replicating the original report by Brand and Deary (1982).

Certain difficulties have emerged from these three studies. Deary (1980) and

Nettelbeck, Edwards, and Vreugdenhil (1986) found that some subjects have problems in making the discrimination regardless of the stimulus duration. Even with Nettelbeck's improved task the range of AITs among university students was 11 to 382 ms with a standard deviation of 101. Irwin (1984) reported that children's AIT scores correlated at  $-0.51$  with scores on the Seashore pitch discrimination test. Further, Irwin demonstrated that at short stimulus durations the overlap in the frequency spectra of the stimulus tones is so great that the AIT task becomes as much a pitch discrimination as a temporal resolution task. This raises the possibility that the AIT-IQ correlation is due to pitch discrimination ability. This suggestion received support from two sources. A recent historical review (Deary, in press) indicates that several studies have obtained pitch discrimination-IQ correlations in the region of 0.2. Second, Raz, Willerman, and Yama (1987) found correlations of between  $-0.42$  and  $-0.54$  among measures of frequency resolution and the Cattell Culture Fair IQ test in college students. Raz and his coworkers suggest that better resolving ability by high IQ subjects may be a determinant of mental speed. This is an important issue which requires consideration. If the AIT-IQ correlation is due to AIT being, in effect, a test of pitch discrimination then the suggestion that general information intake speed is an important component of intelligence loses much of its power.

The present study attempts to resolve these issues. A new form of the AIT test has been devised which yields AIT values which are outside the range where the frequency spectra for the stimulus tones overlap. In pretesting, no undergraduate had an AIT less than 30 ms ( $M$  75.8 ms,  $SD$  27.5,  $n=80$ ; Deary, 1987). If general information processing speed is a component of intelligence, then it is important to know whether this is based upon better stimulus discrimination. Using an improved AIT task the previous studies were extended by testing the following hypothesis: If information intake speed is an important component of intelligence, then AIT will still correlate with IQ scores with pitch discrimination partialled out. Also, we attempted to discover whether AIT has a closer relationship with verbal or with nonverbal ability scores.

## EXPERIMENT 1

### Subjects

Fifty-nine second-year psychology undergraduates were recruited for this study. All subjects were age 19-23; 37 were female. (No sex differences have been found in adult IT studies and sex differences were not examined here.) All subjects had normal or corrected-to-normal vision and had no hearing deficits.

### Methods

*Alice Heim 6 AG* (Heim, Watts, & Simmonds, 1983)—This is an ability test which is able to discriminate among university level adults. It is a 40-minute



group test intended for arts and general university students. It yields two sub-scores, a verbal score and a numerical and diagrammatic score, which may be added to give a total score.

**Seashore Pitch** (Seashore, Lewis, & Saetveit, 1956)—This is a 50-item test where subjects are required to indicate the temporal order (high-low or low-high) of tone pairs which vary in pitch difference from 17 Hz to 2 Hz. The test begins with easy items and progresses to the difficult items. The tones are played for a duration that is well above the AIT of all testable subjects and the tones are not masked. The score is the number of tone pairs discriminated correctly.

**Auditory Inspection Time**—This was constructed as a fixed block of trials that progressed from long (easy) to short (difficult) stimulus durations. Stimuli were square wave tones and were presented using an XR320 monolithic timing circuit driven by a BBC microcomputer. High tones were 880 Hz and low tones were 784 Hz. The auditory mask was a rapidly alternating series of 10 ms bursts of both stimulus tones provided by a multivibrator circuit. Test stimuli were recorded on 1/4 in. tape and played on a UHER reel-to-reel tape recorder. The output was relayed to quiet individual rooms and played through headphones. Sound levels were 80 dB for stimuli and mask.

Each item in the AIT task consisted of a cue tone (832 Hz) lasting 500 ms, 1 s of silence, a stimulus tone pair with no gap between them (high-low or low-high) and 1 s of mask (described above). Thus there was no stimulus-mask gap and no interval between the two stimulus tones.

### Procedure

All subjects were tested on the Seashore test one week prior to the AH6 and the AIT tests. Seashore and AIT testing were done individually in quiet cubicles and the stimuli were played through headphones. AH6 was administered as a group test. Approximately half of the subjects took the AH6 prior to AIT and the others completed these in the reverse order. The AIT task was introduced in general terms to all subjects in a group fashion. Thereafter, detailed instructions and example items were supplied on a prerecorded tape. Subjects were given 8 s between items to record their responses. Twenty items were presented at long durations (200 ms). Ten were masked and ten were unmasked. These served as practice items and were also used to determine whether the subjects were able to perform the discrimination. Experimental trials came in 13 blocks of 10 stimulus pairs (each having 5 high-low and 5 low-high pairs) presented at progressively shorter durations: 200, 150, 125, 100, 85, 70, 55, 40, 30, 20, 15, 10, and 6 ms. Those subjects ( $n = 24$ ) who were unable to achieve 90% correct responses for the mean of the second practice and first experimental block were excluded from further AIT analysis. (We had found in pretesting that this group is not at one end

of the AIT distribution; rather, they appear to be a separate group with poor pitch discrimination skills (Deary, 1987).)

### Results

Table 1 shows the summary data and  $t$ -test results for those subjects who could ( $n = 34$ ) and those who could not ( $n = 25$ ) perform the AIT test. There was no difference in AH6 scores or sub-scores, but the Seashore pitch discrimination scores differed significantly between the two groups ( $P \leq .001$ ). A correlation analysis was performed on the variables for those subjects who could perform the AIT test. Table 2 shows the high correlation between the AH6 subtests. The correlation between pitch discrimination and AH6 is near zero and nonsignificant. The correlation between AIT and AH6 total score is  $-0.39$  ( $p < .01$ , one-tailed), with AH6 verbal score is  $-0.45$  ( $p < .01$ , one-tailed), and with AH6 numerical and diagrammatic score is  $-0.27$  ( $p < .1$ , one-tailed). The correlation between AIT and pitch discrimination was in the expected direction ( $r = -0.20$ ) but nonsignificant. With pitch discrimination scores held constant the partial correlation between AIT and AH6 total score is  $-0.38$ , that is, very little different from the raw correlation.

### Discussion

This study is the first report of the correlation between AIT and IQ-type test scores which takes account of differences in pitch discrimination ability. The

TABLE 1  
Alice Heim 6 and Pitch Discrimination Scores ( $M \pm SD$ ) for Those Undergraduates Who Could and Could Not Perform the AIT Task

	Can do AIT ( $n = 34$ )	Can't do AIT ( $n = 25$ )	$t$	$p$
AH6 verbal	15.8 (4.3)	15.8 (4.9)	.01	ns
AHG N + D	14.2 (4.2)	13.8 (4.7)	.32	ns
AH6 total	30.0 (8.0)	29.6 (8.2)	.18	ns
Seashore pitch	44.3 (2.8)	36.7 (6.0)	5.94	$< .001$
AIT (ms)	77.9 (26.1)			

TABLE 2  
Correlations among Alice Heim 6, AIT and Seashore Pitch Scores for Subjects Able to Perform the AIT Task ( $n = 34$ )

	AH6 Verbal	AH6 N + D	AH6 Total	Seashore Pitch	AIT
AH6 verbal	—	.77***	.94***	-.06	-.45**
AH6 N + D		—	.94***	-.09	-.27
AH6 total			—	-.08	-.39**
Seashore pitch				—	-.20

\*\* $p < .01$ ; \*\*\* $p < .001$

correlation between total score on the AH6 test and AIT was  $-0.39$ , which is very similar to the correlation of  $-0.38$  reported between AIT and Raven's Advanced Progressive Matrices (Nettelbeck, Edwards, & Vreugdenhil, 1986). This study went further and examined the contribution made by pitch discrimination ability in the AIT-IQ correlation. The answer appears to be very little. Our AIT task average scores were well out of the range where frequency spectra overlap is a major problem and the task appears to be a reasonable test of auditory information intake speed. We conclude this because the Seashore-AIT correlation was nonsignificant and the AIT-AH6 correlation decreased very little when pitch discrimination ability is partialled out.

It is noteworthy that 42.4% of the subjects were unable to make the pitch discrimination reliably enough for their data to be included in the AIT analysis. This is in agreement with our previous findings (Deary, 1987), but here those who could and could not make the discrimination were examined in more detail. Those who fail to manage the AIT task were found to: (a) form a separate group on AIT performance, they are not at one end of a normal AIT distribution; (b) have similar AH6 scores to those who can do the AIT task; and (c) have lower pitch discrimination scores than those who can do the AIT task. That pitch discrimination screening of AIT subjects should take place is suggested, just as all visual IT subjects are pretested for their ability to make adequate visual discriminations. Once an adequate level of pitch discrimination is attained it does not appear to play a part in AIT ability; that is, it acts as a threshold on rather than a correlate of AIT. In undergraduates pitch discrimination ability was not related to AH6 scores and we conclude that information intake speed is arguably the key to the AIT-AH6 correlation. By careful elimination of those subjects who were unable to perform the AIT task at any duration, and whose main problem seemed to be one of poor pitch discrimination ability, the problem of large AIT standard deviations reported by Irwin (1984) and Nettelbeck, Edwards, and Vreugdenhil (1986) appears to have been solved. Also we have established, at least by exclusion, that it is the temporal resolution aspect of AIT that relates to IQ test scores. However, to examine further the possibility that any AIT-IQ correlation in children is due to pitch discrimination differences, as Irwin's (1984) findings indicate, a similar exercise to that above was carried out on a large group of children, as follows.

## EXPERIMENT 2

### Subjects

Sixty boys (mean age 12 years 4 months,  $SD$  4.2 months) and 59 girls (mean age 12 years 6 months,  $SD$  5.1 months) took part in this study. All subjects had normal or corrected-to-normal vision and none had hearing difficulties.

### Methods

**Cognitive Ability Tests** — Two verbal ability tests, one nonverbal intelligence test, and one mathematical reasoning test were administered to all subjects. The verbal tests were: the Verbal Reasoning Test from the Moray House series, which yielded an age-corrected verbal reasoning quotient (VRQ); and the Mill Hill Vocabulary Test Form 1 Junior Parts A and B (Raven, Raven, & Court, 1982), which yielded an age-corrected verbal IQ (Mill Hill IQ). The nonverbal test was Raven's Progressive Matrices (Raven, 1958), and this was age-corrected to give a nonverbal IQ (Raven IQ). The mathematical test was the Mathematical Reasoning Test from the Moray House series and it was age-corrected to give a mathematical reasoning quotient (MRQ).

**Pitch Perception Testing** — The pitch discrimination subtest of the Bentley Tests of Musical Ability (Bentley, 1966) was used. The pitch discrimination test is a 20-item test involving tone pairs whose tones are played one after the other with no masking. Stimulus duration of the tones is about 0.5 s and the subject is required to state whether the second tone is higher, lower, or the same as the first tone.

**Auditory Inspection Time (AIT)** — The schoolchildren undertook the same test as the undergraduates in the previous study. The group instructions prior to testing were more extensive and more checks were made to ensure that subjects understood the nature of the test. However, the taped instructions and the test were identical.

### Procedure

The Bentley pitch discrimination test was administered to the subjects in a group fashion in quiet classrooms according to the instructions in the testing manual. The mathematical and verbal reasoning tests were given in classrooms as group tests. Raven, Mill Hill, and AIT tests were administered in the same situations as the previous experiment. All subjects underwent tests in the same order: Raven, Mill Hill, and, after a 30-minute break, a second administration of the Bentley Pitch Test followed by AIT. All ability test raw scores were converted to IQ-type scores and the AIT was scored as in the previous experiment.

### Results

Summary statistics from all tests are presented in Table 3. Twenty-nine of the 60 boys and 24 out of the 59 girls were able to perform the AIT task. This proportion was not significantly different from the 34 out of 59 students who could perform the task ( $\chi^2 = 2.21$ ,  $df = 1$ ). Girls scored significantly better on the pitch discrimination task at the second sitting ( $p < .05$ ) and had higher mathematical

reasoning quotients ( $p < .01$ ); otherwise the two groups were very similar and data analysis was performed on the whole group. Table 3 has the summary data for schoolchildren divided into those who could and those who could not perform the AIT task. There was no age difference between the two groups. Like the student group, pitch discrimination was significantly different between the two groups ( $p < .001$ ). Unlike the student group, both the verbal and nonverbal IQ test scores and the verbal and mathematical reasoning scores were significantly different in the two groups with those able to perform the AIT task having higher mean scores. Therefore, in children, getting started on the AIT task appears to be related to ability level and pitch discrimination ability.

Table 4 shows the Pearson correlations among the ability, Bentley Pitch, and AIT tests. As expected, all ability test intercorrelations are positive and significant. Test-retest reliability on the Bentley Pitch test was 0.52. There are 8 correlations between pitch discrimination ability and mental ability (pitch discrimination estimated at two sittings versus four ability tests) with a range of 0.04 to 0.29 ( $M = 0.15$ ). When these correlations are made to these correlations to take account of the moderate test-retest reliability of the Bentley Pitch test, the range becomes 0.06 to 0.45 ( $M = 0.25$ ). There is a difference between the verbal and nonverbal test correlations with pitch discrimination. Raven and MRQ correlations with pitch discrimination ability are all near to zero and nonsignificant. Three of the four verbal ability test-pitch discrimination correlations are significant and in the expected direction. Therefore, superior pitch discrimination appears to convey an advantage on tests of vocabulary and verbal reasoning that is not apparent in undergraduates. Pitch discrimination scores correlate significantly with AIT at the first pitch test and in the same direction but nonsignificantly at the second testing of pitch.

Correlations between mental ability test scores and AIT are all in the negative direction and are all significant (Table 4). As in the undergraduate sample, it is the verbal IQ which correlates at a higher level with AIT ( $-0.36$ ,  $p < .01$ ). Table

TABLE 3  
Mental Ability Test and Bentley Pitch Scores and Ages ( $M \pm SD$ ) for Those Subjects Who Could and Could Not Perform the AIT Task

	Can do AIT ( $n = 53$ )	Can't do AIT ( $n = 66$ )	$t$	$p$
Raven	116.0 (10.7)	111.6 (10.2)	2.31	<.05
Mill Hill	118.7 (10.4)	113.3 (12.4)	2.50	=.01
VRQ	116.1 (12.4)	108.1 (10.4)	3.81	<.001
MRQ	111.9 (11.2)	106.2 (10.6)	2.81	<.01
Pitch 1	16.9 (1.8)	14.3 (2.7)	6.11	<<.001
Pitch 2	17.1 (2.6)	14.9 (2.8)	4.45	<<.001
Age (months)	140.8 (3.5)	140.9 (5.5)	.06	ns
AIT (ms)	75.7 (31.2)			

TABLE 4  
Correlations among Mental Ability Test, Bentley Pitch and AIT Scores for the Schoolchildren Sample ( $n = 119$  Except for Correlations Involving AIT Where  $N = 53$ )

	Raven	Mill Hill	VRQ	MRQ	Pitch 1	Pitch 2	AIT
Raven	—	.43***	.66***	.64***	.10	.09	-.26*
Mill Hill		—	.55***	.41***	.29**	.22*	-.36**
VRQ			—	.82***	.14	.21*	-.28*
MRQ				—	.04	.12	-.24*
Pitch 1					—	.52***	-.26*
Pitch 2						—	-.11

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ , one-tailed.

4 also shows that the narrower, school performance-related tests of mathematical and verbal reasoning also correlate significantly with AIT scores. Table 5 shows the partial correlations between ability test scores and AIT when pitch discrimination is controlled. The results change very little from the original correlations with the verbal IQ-AIT correlations around  $-0.3$  and the Raven IQ-AIT correlations at about  $-0.25$ .

### Discussion

Here the first replication of the correlation between AIT and IQ in children is reported. However, we have gone further than Irwin's (1984) study. Using an improved AIT test and screening for inability to perform the AIT task at any duration up to 200 ms, even when the stimuli were unmasked, the problems of very long AIT durations and high AIT standard deviations, both of which have introduced a marked contamination of temporal resolution speed with pitch discrimination ability, were eliminated. Pitch discrimination correlated at about 0.15 with IQ in this sample. Partial correlations between IQ and AIT, controlling for pitch discrimination, differ only minimally from the original correlations. It appears, then, that while AIT correlates at low levels with pitch discrimination ability the AIT-IQ correlation is due to temporal resolution speed rather than the ability to make fine pitch discriminations. In children, at least, the abilities of sensory discrimination and temporal resolution appear to be separate correlates of psychometric intelligence.

TABLE 5  
Partial Correlations Between Ability Test Scores and AIT Controlling for Pitch Discrimination ( $n = 53$ )

	Raven	Mill Hill	VRQ	MRQ
Controlling for pitch 1	-.25	-.32	-.27	-.28
Controlling for pitch 2	-.26	-.35	-.28	-.25

Recent research has begun to examine the particular abilities which are related to measures like IT rather than simply replicating IQ-IT correlations. The data from Experiment 2 indicate that the high-verbal child has an advantage on certain sensory indices—he or she has somewhat better pitch discrimination ability and a noticeably shorter AIT. The child scoring highly on nonverbal or mathematical tests appears to have no particular advantage in pitch discrimination and to have less advantage in the AIT task.

This study corroborates the few existing reports which indicate that auditory inspection time (AIT) correlates with cognitive ability. Significant correlations were obtained despite the fact that the students, boys and girls, were of above average ability and, the students in particular, restricted in ability range. More important, the present study has gone some way toward providing an understanding of the AIT-IQ correlation. Previous forms of the AIT test resulted in such low inspection times that the stimulus tones overlapped in frequency. Therefore, previous AIT tests have confounded pitch discrimination ability with information processing speed. By using a more effective mask and by eliminating the inter-tone gap (which allowed rehearsal of the first stimulus tone) the lowest obtained auditory inspection times were increased to about 30 ms without altering the proportion of people able to perform the test. The AIT-IQ correlation in the present study does not rely on outliers in the data. Our standard deviations in the AIT test are lower than those reported by previous workers (Deary, 1980; Irwin, 1984; Nettelbeck, Edwards, & Vreugdenhil, 1986). This is partly due to pretesting for ability to do the AIT test. Subjects were required to attain at least 90% performance on 20 items at long stimulus durations. Subjects who could not do this were found to form a separate group from those who could. Those who did not get started on the AIT test were distinguished, in both the undergraduates and the schoolchildren, by their low pitch discrimination scores. If, as suspected, previous studies have failed to eliminate these subjects from the AIT-IQ analysis then they will have confounded pitch discrimination ability and temporal resolution speed. The fact that over one third of the undergraduates and almost one half of the schoolchildren from the analysis must be excluded is inconvenient but is the equivalent to ensuring that all subjects in, say, the visual test have adequate vision.

It is our opinion that the subjects deemed unable to perform the AIT task were not those at the extreme end of a normal distribution for auditory intake speed. If a subject is having difficulty making the pitch discrimination involved in the AIT task, but is, nevertheless, allowed to complete it, then some kind of AIT score may be attributed to him. However, as Vickers, Nettelbeck, and Willson (1972) indicated, this leads to the introduction of variance from the noise parameter as well as the inspection time parameter in perception. Similarly, if a subject with poor visual acuity is allowed to proceed in the visual IT task with uncorrected vision, his visual IT will be contaminated with noise. In IT testing the stimuli must be easily discriminable in order that the difference between the stimuli is

well beyond the noise level for most subjects. For those subjects included in our AIT-IQ correlations we are confident that this was the case. Their performance on unmasked items and at least the first two blocks of masked AIT items was virtually error free and their falloff in accuracy with decreasing stimulus duration time was, for the most part, regular and progressive.

Those subjects who had they been included in the AIT-IQ analyses, would have had AIT estimates confounded with noise were distinguished from the others on three counts. First, they had a poorer average level of pitch discrimination skills. This index, however, was not a perfect discriminator as there was a slight overlap in the pitch discrimination distributions of those who could and those who could not perform the AIT task. Second, those who could not do the task were the only subjects to make regular errors in unmasked practice trials. Third, they showed no regular falloff in accuracy with progressive reductions in duration; their accuracy levels were often slightly above chance for the longer durations, but showed no psychometric curve. It should be emphasized, however, that the important point is that some subjects were excluded before IQ results were known to prevent the IT parameter being contaminated with the noise parameter, as was believed to have been the case in other studies. Therefore, the present study is the first to offer an unequivocal result which links auditory information intake speed and intelligence and which does not confound pitch discrimination ability.

In undergraduates pitch discrimination ability was found not to correlate with intelligence and pitch discrimination was found to be the only variable which distinguished those subjects who could not perform the AIT test. Also, the correlation between AIT and pitch discrimination was nonsignificant. The schoolchildren had very similar mean AITs when compared with the undergraduates but a smaller percentage of them were able to perform the AIT test. It appears that in undergraduates pitch discrimination acts in a threshold manner such that above a critical level of pitch discrimination ability there is very little or no additional advantage in performing AIT. Also, what small overlap exists between AIT and pitch discrimination does not coincide with the variance shared between AIT and intelligence. In children it appears that intelligence is a stronger correlate of pitch discrimination. Also, those subjects who could not perform the AIT were distinguished by their lower IQs as well as their poorer pitch discrimination ability.

This study offers evidence from two separate samples with different ages about "what it means to be high-verbal" (Hunt, Lunneborg, & Lewis, 1975). This study indicates that the AIT-IQ correlation is higher in tests of verbal IQ rather than in nonverbal or mathematical tests. Also, the high-verbal schoolchildren had a pitch discrimination advantage, but no such pitch discrimination-verbal ability correlation was found in the undergraduate sample. It appears, then, that advantages in two simple auditory abilities—those of auditory processing speed and of pitch discrimination—correlate particularly well with verbal



ability scores. Speculatively, these advantages may have an ontogenetic relationship with the verbal ability scores, if faster information intake and more accurate discrimination allow more resources to be freed for the consolidation of verbal information. In Vickers' terms we hypothesize that a child with AIT and pitch discrimination advantages need make fewer 'inspections' of a stimulus before an accurate decision is made. This has the effect of reducing the time and effort spent on each discrimination and allows the subject to move on to another stimulus more quickly, or gives the subject more time to rehearse the already discriminated stimulus. This hypothesis, however, does not explain why the verbal ability-AIT correlations were higher than those between nonverbal ability and AIT. Perhaps the distinction between the process and the products of intelligence may be invoked here. The products of intelligence, as indexed by verbal ability, are the accumulation of many instances of the process in operation. As such, a product test may deliver a reliable average of past performances of the process of intelligence. The process of intelligence, as indexed by tasks like Raven's Progressive Matrices, is measured at a single sitting and is more liable to be unrepresentative of average function. Thus it may be that the AIT-verbal ability correlation is higher because the verbal ability score is acting as a cumulative average of past levels of processing efficiency and that it has less idiosyncratic variance than a more 'fluid' task.

Because, from our previous work (Deary, 1987), we were aware that the undergraduates have an average IQ of about 124, we did not intend to compare the groups in experiments 1 and 2 since they differ in age and intelligence. However, they differed very little in their AIT levels. This finding must be added to the debate about whether and over what range of chronological ages IT develops (Anderson, 1986; Brand, 1984). In fact, the present study is rare in presenting the same IT test to children and adults and in this first study of AIT in different ages processing speed is found not to improve from age 12.5 to 21. This is an interesting result which requires further investigation since it indicates that, while  $g$  increases up to the late teens, IT may be constant from an earlier age. In conclusion, the AIT-IQ correlation is due to AIT being an index of information intake speed, not of pitch discrimination ability. AIT is found to correlate with different tests of mental ability, and especially well with verbal ability. AIT does not appear to improve from age 12.5 to age 21.

## REFERENCES

- Anderson, M. (1986). Inspection time and IQ in young children. *Personality and Individual Differences*, 7, 677-686.
- Bentley, A. (1966). *Musical ability in children and its measurement*. London: Harrap.
- Blinkhorn, S.F., & Hendrickson, D.E. (1982). Average evoked potentials and psychometric intelligence. *Nature*, 295, 596-597.
- Brand, C.R. (1984). Intelligence and inspection time: an ontogenetic relationship? In C.J. Turner & H.B. Miles (Eds.), *The biology of human intelligence*. Humberston, England: Nafferton.
- Brand, C.R., & Deary, I. (1982). Intelligence and 'inspection time'. In H.J. Eysenck (Ed.), *A model for intelligence*. New York: Springer.
- Deary, I.J. (1980). *How general is the mental speed factor in 'general' intelligence? An attempt to extend inspection time to the auditory modality*. Unpublished B.Sc. Honours Thesis, University of Edinburgh, Scotland.
- Deary, I.J. (1987, June). *Auditory and visual inspection time as group tests: their intercorrelation and relationship to IQ scores*. Paper presented at the third biennial meeting of the International Society for the Study of Individual Differences, Toronto, Canada.
- Deary, I.J. (in press). The nature of intelligence: simplicity to complexity and back again. In D. Forsshaw & M. Shepherd (Eds.), *The Maudsley essay series in the history of psychiatry and psychology*. London: Vade Mecum.
- Heim, A.W., Watts, K.P., & Simmonds, V. (1983). *AH6 Group Tests of High Level Intelligence Manual*. Windsor, England: NFER-Nelson.
- Hunt, E.B., Lunneborg, C., & Lewis, J. (1975). What does it mean to be high verbal? *Cognitive Psychology*, 7, 194-227.
- Irwin, R.J. (1984). Inspection time and its relation to intelligence. *Intelligence*, 8, 47-65.
- Jensen, A.R., & Vernon, P.A. (1986). Jensen's reaction-time studies: a reply to Longstreth. *Intelligence*, 10, 153-179.
- Mackenzie, B., & Bingham, E. (1985). IQ, inspection time and response strategies in a university population. *Australian Journal of Psychology*, 37, 257-268.
- Mackenzie, B., & Cumming, S. (1986). How fragile is the relationship between inspection time and intelligence: the effects of apparent motion cues and previous experience. *Personality and Individual Differences*, 7, 721-729.
- Mackintosh, N.J. (1986). The biology of intelligence? *British Journal of Psychology*, 77, 1-18.
- Nettelbeck, T. (1987). Inspection time and intelligence. In P.A. Vernon (Ed.), *Speed of information processing and intelligence*. Norwood, NJ: Ablex.
- Nettelbeck, T., Edwards, C., & Vreugdenhil, A. (1986). Inspection time and IQ: evidence for a mental speed-ability association. *Personality and Individual Differences*, 7, 633-642.
- Raven, J.C. (1958). *Standard Progressive Matrices*. London: Lewis.
- Raven, J.C., Raven, J., & Court, J.H. (1982). *Mill Hill Vocabulary Scale*. London: Lewis.
- Raz, N., Willerman, L., & Yama, M. (1987). On sense and senses: intelligence and auditory information processing. *Personality and Individual Differences*, 8, 201-210.
- Seashore, C.E., Lewis, D., & Saeveit, J.G. (1956). *Seashore Measures of Musical Talents*. New York: Psychological Corporation.
- Taylor, M.M., & Creelman, C.D. (1967). Efficient estimate on probability functions. *Journal of the Acoustical Society of America*, 41, 782-787.
- Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Personality and Individual Differences*, 7, 609-623.
- Vickers, T., Nettelbeck, T., & Willson, R. J. (1972). Perceptual indices of performance: the measurement of 'inspection time' and 'noise' in the visual system. *Perception*, 1, 263-295.

## AUDITORY INSPECTION TIME: REVIEW, RECENT RESEARCH AND REASSESSMENT

Ian J. Deary

University of Edinburgh

This study examined the hypothesis that individual differences in auditory inspection time (AIT) are largely due to differences in pitch discrimination ability. We tested 120 schoolchildren (60 boys, 60 girls) of mean age 11.5 years on two verbal and non-verbal ability tests and on the Bentley tests of Musical Ability. Subjects were also tested on a novel form of the AIT test which was devised in response to the criticisms of an earlier version of the test (Irwin, 1984). About 50 % of schoolchildren (compared with 67 % of undergraduates) were able to perform the AIT task. In the boy's group, all ability tests correlated at about -.35 with AIT durations. In the girl's group, AIT durations and verbal ability tests correlated at about -.4, while non-verbal tests showed much lower correlations with AIT. Pitch perception ability correlated at about -.2 with all ability tests and with AIT. Partial correlations between AIT and ability tests, controlling for pitch perception, achieve levels of about -.3. Factor analysis reveals that AIT variance shared with pitch perception may be attributed to g.

### 1. INTRODUCTION

The perceptual index known as 'inspection time' (IT) was developed in the early 1970s as one aspect of an accumulator model of visual perception (Vickers, Nettelbeck and Willson, 1972). An individual's IT is the minimum stimulus presentation time that allows him to reach a predetermined level of accuracy in a decision task. The decision task is usually simple, and it often involves deciding which of two lines of markedly different length is longer. Any one of a number of psychophysical procedures (e.g., adaptive staircase, method of limits, or method of constant stimuli) may reliably be used to estimate the duration required by a subject in order to make, say, 85% correct decisions in a two choice task. IT must be distinguished from reaction time: its measurement does not involve the subject responding rapidly. In the estimation procedure only the correctness of the responses is recorded;



subjects respond at leisure and accuracy is stressed over speed of response.

In the eleven years after Nettelbeck and Lally (1976) first reported a significant correlation between IT and psychometric measures of intelligence, there were over 20 attempted replications of the result. Others have reviewed these studies (Brand and Deary, 1982; Vernon, 1986; Nettelbeck, 1987), and I shall not replicate these efforts here. Briefly, the mean uncorrected correlation between IT and IQ-type scores in all studies to date is  $-.35$ . However, these studies include a preponderance of samples where the mean IQ is high and the range of ability is restricted (because undergraduates are a convenient source of subjects), and the mean disattenuated IT-IQ correlation is about  $-.50$  (Nettelbeck, 1987). The relationship between IT and IQ holds for both verbal and non-verbal tests of intelligence; and it holds in many different sample types, including the mentally handicapped, young children, the elderly and college students. Given that a moderate relationship between IT and IQ exists, the main focus of recent efforts has been to discover the reason for this correlation.

There are two broad views with regard to explanation: IT is either a cause of or a consequence of high intelligence. Take the latter position first. It may be argued that there are aspects of the visual IT task that make it easier for those with high IQs. Some have investigated the possibility that successful completion of IT tasks involves using strategies to penetrate apparent movement cues which result in artefactual fast ITs (MacKenzie and Bingham, 1985; MacKenzie and Cumming, 1986). Another 'consequence' hypothesis is that high IQ subjects are more able and/or willing to comply with boring tasks like IT, RT and evoked potential stimuli (Mackintosh, 1986) and to maintain attention during them. Both of these views interpret IT as an uninteresting consequence of high IQ and, if either is correct, then IT cannot be offered as the basis for a substantial part of IQ variance.

The second hypothesis has it that IT is an index of mental speed which may explain some of the variance in IQ-type test scores. Brand (1984), Brand and Deary (1982) and Deary (in press) have argued that a fast inspection time is ontogenetically related to high levels of intelligence: the quicker completion of processing any one discriminandum leaves the subject more time to make more discriminations or enables him to make more detailed discriminations of the same stimulus, and this leads to the building up of greater levels of stored knowledge. The nature of this mental speed has been debated. Brand and Deary (1982) tended to interpret the IT as an index of general mental speed, perhaps reflecting a general neural efficiency or fidelity of information transfer, while others (see Nettelbeck, 1987) have

hypothesised that IT reflects the speed of information input processes, particularly that stage of processing where information is passed from sensory registers to short term memory.

This is the background of empirical evidence and theory which drives current experimental work on IT. In Edinburgh, we have investigated IT along four fronts. We have examined the brain potentials which are evoked by the IT stimuli and we have found some reliable correlations among IT, IQ and AEP indices (Zhang, Caryl and Deary, *in press*). Second, we are currently investigating the effects of blocking or enhancing various neurotransmitter pathways on the performance of IT tasks (Petrie, 1988). Third, we have attempted to test various hypotheses put forward by those who see strategies as the key to performing the IT task (Egan, 1986). Fourth, we have developed an auditory version of the IT task in order to extend the IT studies into a modality other than vision.

It is the auditory IT work which will be discussed in detail in this paper. The reason for devising the auditory IT (AIT) task was that, if the visual task was more than a modality-specific perceptual trick, and if IT was a measure of general mental speed, then two hypotheses follow. First, AIT should correlate with IQ. Second, AIT should correlate with visual IT (VIT).

## 2. EARLY AUDITORY INSPECTION TIME STUDIES

Deary (1980, reported in Brand and Deary, 1982) devised an AIT task with the intention of indexing mental speed in a way that was analogous with VIT. The auditory discrimination was simple: which of two square wave tones of markedly different pitch, and presented one after the other, was presented first. Subjects were alerted by a cue, heard the first tone (either 770 or 880 Hz), heard white noise for 500 ms and then heard the second tone (880 or 770 Hz). Each tone was played for the same duration, ranging from 200ms, to 2.7ms. Both tones were forward and backward masked using white noise. The test was presented as blocks of trials, beginning with blocks of stimuli of long duration and proceeding to shorter and more difficult durations. AIT represented that part of the performance curve where subjects were 90% correct in their judgements regarding the order of the stimuli. AIT correlated at  $-.70$  with Raven's Progressive Matrices and at  $-.66$  with Mill Hill Vocabulary scores ( $n=13$ ; two subjects were mentally handicapped). In Deary's (1980) study AIT and VIT were correlated at a level near to unity, but the correlation was dependent upon the inclusion of the mentally handicapped subjects.

Two independent attempts to replicate this result followed in the next six years. Irwin (1984) used a task very similar to that of Deary

(1980). He tested 50 twelve year old children and obtained correlations between the AIT, the Raven's Matrices and the Mill Hill Vocabulary scores of  $-.23$  and  $-.32$ , respectively. While these were significant, Irwin failed to replicate the significant AIT-VIT correlation. Nettelbeck, Edwards and Vreugdenhil (1986) improved the AIT task by replacing the white noise mask with a mask consisting of 15ms alternating bursts of both target tones. Their new mask was used before, between and after the tones, for 1000ms in each case. IQ was tested in 30 subjects using Raven's Advanced Progressive Matrices and the AIT-IQ correlation was  $-.38$ , while the VIT-AIT correlation was  $.39$ .

### 3. PROBLEMS WITH, AND IMPROVEMENTS IN, THE AIT TASK

Although the AIT task had the advantage of not being penetrable by strategies, which remain a possibility with those VIT stimuli which produce apparent movement effects, the AIT task, as devised by Deary (1980) and developed by Nettelbeck et al. (1986), had some obvious problems which did not occur with the VIT task. First, the AIT testing tended to result in a skewed distribution of scores, with most subjects obtaining very short AITs and a few obtaining very long AITs. Typical of the results are those of Deary (1980), where the range of AITs was 6 to 160ms, but where the median was 10ms. There were at least three factors contributing to this type of distribution. First, the inclusion of mentally handicapped subjects tended to result in those subjects obtaining very long AITs. Second, white noise was an ineffective mask and allowed subjects to continue to extract stimulus information from the sensory representation of the stimulus, even after the mask had begun. Third, although subjects were being pre-tested for pitch discrimination ability, those subjects who found the basic discrimination in the AIT task very difficult were not always being left out of the later analyses. This amounts to something like allowing subjects with very poor visual acuity to proceed with the visual IT task: it introduces pitch discrimination variance into a task which is supposed to be measuring speed of processing only. A second problem with the original AIT task was the overlap in frequency spectra at very short tone durations. Irwin (1984) provided evidence which demonstrated that, at durations of about 10ms (a region where many subjects were able to make correct AIT discriminations), the target tones had large overlaps in their frequency spectra, a phenomenon not found when the target tones were played for, say, 75ms. Thus, as the duration of the tone pairs became shorter, the AIT task might have been an amalgam of both a perceptual speed and a pitch discrimination task. Finally, the AIT task was not analogous with the VIT task, because it allowed the subject some time between the two stimuli: subjects were able to rehearse the first of the two tones in short term memory in the 500 or 1000ms between the tones, and the task may have allowed rehearsal of

the second tone because the mask was not effective.

In view of these problems, Deary, Caryl, Egan and Wight (in press) devised a new auditory IT task. As before, subjects made a decision about the order of two tones which were of markedly different pitch (880 and 784 Hz). The mask came on immediately after the second tone had ended, and took the form of a series of alternating 10ms bursts of the two target tones. There was no space between the two target tones. Subjects were trained on the new task, and those who were unable to make reliable decisions at long tone durations were excluded from further analysis. (In fact, we allowed these subjects to complete all blocks of AIT trials, but we found that they formed a group whose scores had no overlap with those subjects who were able to respond reliably to longer durations on the task.) This task appears to have ended the problem of very short AITs: our first testing on 80 undergraduates resulted in a mean AIT of 75.8ms (SD 27.5) and no individual had an AIT of less than 30ms. The same mean level and distribution of AIT durations was found in the studies which follow.

In the first study using this new AIT task, we tested various subgroups of 120 undergraduates on AIT, the three most commonly used versions of the VIT task (vertical lines, horizontal lines, and the stimuli used by Longstreth et al. (1986)), the Alice Heim 5 IQ test, Mill Hill Vocabulary and Raven's Advanced Progressive Matrices (Deary et al., in press). With the exception of the AH5, the standard deviations on the IQ tests, were very small. The AH5-IT correlations, for the three visual tasks, ranged from -.29 to -.33. The AH5-AIT correlation was -.31. Correlations between AIT and the three VIT tasks ranged from .24 to .53. We have replicated the AIT-IQ correlation using this new task on two further samples (Deary, Head and Egan, in press). In a sample of 34 undergraduates the Alice Heim 6-AIT correlation was -.39 (the AIT and the verbal IQ component of the AH6 correlation was -.45); and in a sample of 53 twelve year old schoolchildren the correlations between AIT and Mill Hill Vocabulary and Raven's Matrices scores were -.36 and -.26 respectively.

#### 4. SENSORY DISCRIMINATION OR PROCESSING SPEED?

While the new AIT task has taken the AIT durations away from the region where pitch discrimination probably confounds duration as the key variable, the question as to whether some AIT variance is attributable to individual differences in pitch discrimination remains to be addressed directly. Irwin (1984) found that AITs correlated at -.51 with scores on the Seashore test of pitch discrimination. But, recall that Irwin was using the white noise-masked task, and that he did not report excluding those



subjects unable reliably to perform the AIT task at long durations before analysing his data. Deary et al. (in press) included the Seashore test for the undergraduate sample and tested the schoolchildren sample twice on the shorter Bentley pitch discrimination test. In the undergraduate sample pitch discrimination did not correlate significantly with either IQ measures or AIT, and when pitch discrimination ability was partialled out of the AIT-IQ correlation there was almost no change. In the schoolchildren, pitch discrimination correlated at about  $-.2$  with a test of verbal ability, but at non-significant levels with Raven's IQ and mathematical ability. Also in schoolchildren, pitch discrimination correlated at  $-.18$  with AIT, but, when pitch discrimination was partialled out of the AIT-IQ correlations, they remained almost the same as before.

From our studies, then, it appears that pitch discrimination does correlate at low levels with both AIT and verbal IQ in children, but not in high ability adults. In fact, there are several reports in the literature where measures of pitch discrimination correlate with measures of intelligence, especially in schoolchildren (this was the subject of a recent historical review and reanalysis by Deary (in press b)).

Interestingly, there has been a similar discussion of the interplay between processing speed and sensory discrimination in a separate series of studies. Raz, et al., (1983) and Raz and Willerman (1985) used a backward recognition masking test to estimate undergraduates' speed of auditory processing. This is done by playing a single target tone, either 770 or 870 Hz, for 20ms, leaving a variable interval between the tone and a mask, and masking with a tone which is intermediate in pitch between the two possible targets. Subjects indicate whether the target was high or low. Tone to mask interval may take one of the following values: 0, 20, 30, 60, 120 or 480ms. The number of correct identifications of target tones may be seen as a measure of auditory processing speed, and the correlation between ability on this task and IQ scores ranged from  $-.37$  to  $-.49$  in their 1983 study and from  $-.37$  to  $-.53$  in the 1985 study ( $n=36$ ).

These results may be interpreted as straightforward corroboration of the AIT-IQ correlations, but a more recent study by Raz, Willerman and Yama (1987) raises the issue of pitch perception once again. Using an adaptive staircase procedure, they played subjects two 20ms tones which were 850ms apart and unmasked. Subjects were asked to indicate whether the high tone came first or second. The algorithm sought the smallest pitch difference where subjects could make accurate discriminations. At the outset, the two tones were 100Hz apart and, in some cases, this was reduced to 2 to 3 Hz. In their first experiment, IQ was correlated with frequency discrimination ability for two signal ramps at  $-.42$  and  $-.54$  ( $n=25$ ). The correlation was not caused by individual differences in practice, stimulus spectral composition, musical experience or

demographic characteristics. In a second sample, using undergraduates pre-selected to provide a wide range of IQs, IQ correlated at  $-.50$  and  $-.52$  with two discrimination indices. Again, the correlations were almost unchanged when musical experience was partialled out. In their third experiment, Raz, Willerman and Yama (1987) devised a signal detection task which was just as complicated as the previous discrimination task, but found no correlation between signal detection threshold and cognitive ability. They concluded that, while signal recognition correlates with ability, signal detection does not.

It is also interesting to note that the third experiment of Raz, Willerman and Yama (1987) was designed to test the hypothesis that the high IQ subject simply does better on any novel 'non-entrenched' task, and that it is this fast adaptation to the strange situation in the laboratory that explains the typical IQ-IT correlation. This, they argued, was a suggestion put forward by many cognitive psychologists, and it was refuted by their findings. However, note that this hypothesis is almost the opposite of that thought up by MacIntosh (1986), who suggests that the high IQ subject performs better given the boring nature of laboratory tasks, such as IT. It appears that attempts to explain away the IT-IQ correlation in cognitive terms can be extremely flexible.

## 5. CONCLUSIONS

If the many VIT studies, the AIT studies mentioned here and the studies of Raz and his co-workers are taken together we may make some tentative conclusions. There is much evidence to indicate that both visual and auditory processing speed correlate at moderate levels with psychometric measures of intelligence. Also, in three out of four studies, VIT and AIT have a moderate intercorrelation. Our recent studies appear to indicate that AIT is more closely related to verbal IQ than to non-verbal IQ, indicating that processing speed advantages in one modality may provide the basis for advantages in specific types of cognitive ability. The moderate intercorrelations of various forms of the VIT task (Deary, et al., (in press)) suggest that each task has a degree of task-specific variance, as well as general processing speed variance. Our evoked potential studies suggest that these two sources of variance have separate correlates. The general speed factor appears to be related to IQ, whereas the more task-specific variance is correlated with evoked potential indices of the early stages of information intake or pattern recognition (P200 rise time particularly) (Deary and Caryl, 1988; Hall, 1988).

Although our studies appear to have ruled out the involvement of pitch discrimination as a confounding variable in the AIT task, the results of Raz, Willerman and Yama (1987) indicate that, with briefly presented



auditory stimuli, high IQ subjects make finer pitch discrimination judgements. This should make us wary of prematurely concluding that speed of processing is the key variable, even in tasks which appear to index speed of processing only. Raz and his colleagues (1987) argue that the IT-IQ correlations may be explained by the high IQ subject having: faster feature extraction; better sensory representation of stimuli; faster decision time; less bias in responding; or a combination of the above. The decision time and response bias differences were ruled out by Nettelbeck and Lally (1976) and Lally and Nettelbeck (1977). The remaining novel suggestion, therefore, is that the high IQ subject has a better representation of sensory stimuli. Raz and colleagues remind us that, "quality of signal representation rather than speed of processing may be the key feature of an intelligent brain." However, speed of processing and quality of representations are probably neither alternatives nor explanations at the same level.

If we go back to the theory of Vickers, et al. (1972), it states that individuals, when making decisions about a stimulus in a two-choice decision task, must accumulate evidence for the two options against a background of noise. Evidence accumulates by the subject making inspections of the stimulus, each inspection takes a minimum amount of time, and a decision is made when the evidence for one of the alternatives passes some threshold. In the IT task, the stimulus presentation time is manipulated and, at very brief durations, the subject has not been able to accumulate sufficient evidence from the stimulus to make reliably correct decisions. However, for any given stimulus duration, because of individual differences in IT, some subjects will make more inspections than others, and a subject with a very short IT will achieve a faithful representation of a brief stimulus, whereas a subject with a long IT will have a poor representation. Therefore, a fast IT may cause better representation of stimuli, and Raz's finding that high IQ subjects make better pitch discriminations to briefly presented tone pairs may be explained by the fact that, given a constant presentation time, as the pitch discrimination becomes more difficult, more inspections of the stimuli need to be taken in order to make reliable discriminations. This is unlikely to be a factor in the Seashore test, where the stimuli are presented for 500ms (which is far longer than any AITs reported in the above studies) and are unmasked. However, with 20ms tones of very similar pitch, even when they are unmasked and separated by 850ms, it is likely that the subject with superior IT will have an advantage.

In summary, the argument which pits processing speed against fidelity of stimulus representation may well be a non-argument: the experiments reported here are congruent with an explanation which states that a fast inspection time is primarily an advantage in information processing speed, and that this results in more faithful representation of

briefly presented stimuli.

## CORRESPONDENCE

Correspondence concerning this paper should be addressed to Dr Ian Deary, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ U.K.

## REFERENCES

- Brand, C.R. (1984) Intelligence and inspection time: an ontogenetic relationship? In C.J. Turner and H.B. Miles (Eds), *The Biology of Human Intelligence*. Nafferton, North Humberside: Nafferton Books.
- Brand, C.R. and Deary, I.J. (1982) Intelligence and 'inspection time'. In H.J. Eysenck (Ed.), *A Model for Intelligence*. New York:Springer Verlag.
- Deary, I.J. (1980) How general is the mental speed factor in 'general' intelligence: an attempt to extend inspection time to the auditory modality. B.Sc.(Hons) Thesis, University of Edinburgh.
- Deary, I.J. Intelligence and encoding speed in infants, adults and children. C.P.C. : *Cahiers de Psychologie Cognitive / European Bulletin of Cognitive Psychology*, 8, 462-468.
- Deary, I.J. The nature of intelligence: simplicity to complexity and back again. In D. Forshaw and M. Shepherd (Eds), *Maudsley Essay Series in the History of Psychiatry and Psychology* (in press).
- Deary, I.J. and Caryl, P.G. (1988) Inspection time and cognitive ability. Paper presented at the International Conference on Thinking, Aberdeen.
- Deary, I.J., Caryl, P.G., Egan, V. and Wight, D. Visual and auditory inspection time: their interrelationship and correlation with intelligence in high ability subjects. *Personality and Individual Differences*, (in press).
- Deary, I.J., Head, B. and Egan, V. Auditory inspection time, intelligence and pitch perception. *Intelligence*, in press.
- Egan, V. (1986) Intelligence and inspection time: do high-IQ subjects use cognitive strategies? *Personality and Individual Differences*, 7, 695-700.
- Hall, B. (1988) Evoked potentials, auditory inspection time and intelligence. M.A.(Hons) Thesis, University of Edinburgh.

- Irwin, R.J. (1984) Inspection time and its relation to intelligence. *Intelligence*, 8, 47-65.
- Lally, M. and Nettelbeck, T. (1977) Intelligence, reaction time and inspection time. *American Journal of Mental Deficiency*, 82, 273-281.
- Longstreth, L.E., Walsh, D.A., Alcorn, M.B., Szeszulski, P.A. and Manis, F.R. (1986) Backward masking, IQ, SAT and reaction time: interrelationships and theory. *Personality and Individual Differences*, 7, 643-652.
- MacIntosh, N.J. (1986) The biology of intelligence? *British Journal of Psychology*, 77, 1-18.
- Mackenzie, B. and Bingham, E. (1985) IQ, inspection time and response strategies in a university population. *Australian Journal of Psychology*, 7, 257-268.
- Mackenzie, B. and Cumming, S. (1986) How fragile is the relationship between inspection time and intelligence: the effects of apparent motion cues and previous experience. *Personality and Individual Differences*, 7, 721-729.
- Nettelbeck, T. (1987) Inspection time and intelligence. In P.A. Vernon (Ed.), *Speed of Information Processing and Intelligence*. New Jersey: Ablex.
- Nettelbeck, T., Edwards, C. and Vreugdenhil, A. (1986) Inspection time and IQ: evidence for a mental speed-ability association. *Personality and Individual Differences*, 7, 633-642.
- Nettelbeck, T. and Lally, M. (1976) Inspection time and measured intelligence. *British Journal of Psychology*, 67, 17-22.
- Petrie, R. (1988) The effect of smoking on human information processing. B.Sc.(Hons) Thesis, University of Edinburgh.
- Raz, N. and Willerman, L. (1985) Aptitude-related differences in auditory information processing: effects of selective attention and tone duration. *Personality and Individual Differences*, 6, 299-304.
- Raz, N., Willerman, L., Ingmundson, P. and Hanlon, M. (1983) Aptitude-related differences in auditory recognition masking. *Intelligence*, 7, 71-90.
- Raz, N., Willerman, L. and Yama, M. (1987) On sense and senses: intelligence and auditory information processing. *Personality and Individual Differences*, 8, 201-210.
- Vernon, P.A. (1986) Inspection time: does it measure intelligence? *Personality and Individual Differences*, 7, 715-720.

- Vickers, D., Nettelbeck, T. and Willson, R.J. (1972) Perceptual indices of performance: the measurement of 'inspection time' and 'noise' in the visual system. *Perception*, 1, 263-295.
- Zhang, Y., Caryl, P.G. and Deary, I.J. Evoked potential correlates of inspection time. *Personality and Individual Differences* (in press).

## Inspection Time and Cognitive Ability

IAN. J. DEARY AND PETER G. CARYL

*University of Edinburgh*

### INTRODUCTION

The search for a basic information processing measure that would explain a substantial proportion of the between-individual variance in psychometric intelligence test scores will provide future historians of psychology with a saga whose dates span the twentieth century. Hunt (1980) compared the effort with the search for the holy grail, and it is easy to see why: from glorious, if misinterpreted, failures (Wissler, 1901), through passed-over clues (McK. Cattell, 1886; Burt 1909/10; and see Deary, 1986) and partly justified hubris (Brand and Deary, 1982) the tale has enough scandal, folly, disappointment and, perhaps, achievement to match King Arthur's escapades. But, myth and romanticism aside, there are tangible results from the hypothesis, held by many psychologists since their subject began and by the layman (Sternberg *et al.*, 1981), that some form of mental speed is related, if not causal, to intelligence.

Most fanciers of this hypothesis have reached for their reaction time (RT) apparatus. Why they have done so in such numbers is not clear. It may have been nothing more complicated than the sheer convenience of having RT devices to hand, as opposed to having to invent new *mental* speed tests. Nevertheless, there is sufficient evidence to show that the correlation between intelligence test scores and RT indices will account for a small but significant proportion of variance (Beck, 1933; Jensen and Vernon, 1986). As early as 1890, McK. Cattell warned that RT involved too much variance attributable to motor processes and that a more mental index should be used.

It is ironic that, while he was still at Leipzig with Wundt, Cattell (1886) had devised a reaction-free mental speed task and had already proposed that 'perception time' might have an ontogenic relationship with later mental ability. In parallel with the significant but meagre IQ-RT correlations, there is a history

of much stronger correlations between IQ-type test scores and measures of perceptual intake speed, where the dependent variable is the stimulus presentation time needed by a subject in order to make a correct decision, not the time it takes to react to a standard stimulus (see Deary, 1986, for a review). However, these studies did not form a united body of research (most of the authors seemed unaware of their predecessors) and were not developed further to provide a theory of how the mental speed-intelligence relationship came about.

The inspection time measure did not appear on the scene as yet another arbitrary performance index to be unthinkingly correlated with IQ test scores; it emerged from a psychophysical theory developed by Vickers, Nettelbeck and Willson (1972) in order to give a statistical account of how individuals sample information from visual stimuli when making a decision in a two-choice discrimination task. They proposed that information from the stimulus is sampled in quanta, through the execution of a number of 'inspections'. Reliably correct discriminations are made when a subject has had sufficient time to perceive the stimulus. 'Inspection time' is the name given to the time (in ms) that is required by an individual, under given standard conditions, to make an accurate (usually, but arbitrarily, set at about the 85% correct point on the psychometric curve) decision concerning a simple stimulus. The theory states that, if a stimulus is simple enough, it may require only a single inspection to make an accurate discrimination.

Vickers, Nettelbeck and Willson (1972) hypothesized that IT was a measurable parameter of an individual's thinking processes, reflecting stable individual differences. To date, the most popular form of the IT test uses a stimulus consisting of two parallel vertical lines of markedly different lengths. The lines are presented via a tachistoscope or a LED display and are immediately backward masked. Subjects are required to indicate whether the longer of the lines was on the right or on the left. Responses are made at leisure and, usually, no RT is recorded, only the correctness of the judgement. Stimulus time is varied from the very easy to the impossible, including enough points in between to map the subject's psychometric curve. Subjects' ITs are estimated using various psychophysical techniques, including adaptive algorithms and methods of constant stimuli and, over more than a dozen studies, the IT test-retest reliability is greater than 0.7 (Nettelbeck, 1987).

Interest in, and controversy surrounding, IT grew when it was reported to have a substantial correlation with psychometric measures of intelligence (Nettelbeck and Lally, 1976; Brand and Deary, 1982; Mackintosh, 1981; Nettelbeck, 1983). Two near-complete reviews of the IQ-IT literature (Brand and Deary, 1982; Nettelbeck, 1987) concluded that there is a moderate correlation between these two variables, and that it is probably in the region of  $-0.5$  (that is, higher-IQ subjects have shorter ITs). The correlation between IT and cognitive ability holds for both verbal and non-verbal ('culture-fair') tests and in many different subject populations (normal and mentally handicapped adults, children, the elderly and undergraduates).

There have been various attempts to explain the correlation between cognitive ability and IT. Some have tested the possibility that the correlation occurs because high-IQ subjects adopt task-specific strategies (Mackenzie and Bing-



ham, 1985; Mackenzie and Cumming, 1986), while others (for example, Mackintosh, 1986) have suggested that high-IQ subjects remain more attentive during, and become less bored by, the IT task. If a fast IT involves little more than adopting a particular visual strategy in (or maintaining concentration in) a repetitive and artificial task, then its value as an explanation of human thinking efficiency is limited. Here, IT is seen as a consequence of intelligence.

Others have argued that the IT-IQ correlation arises because IT is a basic feature or component of human information processing or that it might represent neural efficiency (Deary, 1988; Nettelbeck, 1987; Jensen, 1985). In this view IT may be seen as causal to measured intelligence: advantages in the early stages of information encoding, or in general neural information transfer, are said to lead to more accurate and quickly constructed internal representations and to higher levels of fluid and crystallized intelligence.

We shall now report the results of two sets of experiments carried out recently in Edinburgh. In the first section we deal with results from experiments on an auditory version of the IT task. It was felt that if the IT-IQ result could not be replicated using stimuli in another modality, then those who suspected that IT might be a task-specific skill would remain to be answered. In the second section we consider the results obtained when subjects' cerebral evoked potentials are recorded while they perform IT tasks. These experiments were undertaken in the belief that, if there are evoked potential parameters which correlate with IT performance, then we may be able to say something about its importance and place in human information processing.

### AUDITORY INSPECTION TIME

While Nettelbeck (1987) was in a position to review more than two dozen studies examining the relationship between intelligence and visual IT, there are only six studies on the auditory IT-IQ correlation. In various permutations these have addressed one or more of the following questions:

- a) Do auditory IT and psychometric intelligence test scores correlate significantly?
- b) Is there a significant correlation between visual and auditory IT?
- c) Is any auditory IT-IQ correlation which exists due to individual differences in pitch discrimination ability?

Deary (1980; reported in Brand and Deary, 1982) devised an auditory IT task intended to tap auditory processing speed in a manner analogous to the visual task. Instead of the lengths of lines, the pitches of two tones (of 880 and 770 Hz) were chosen as the discriminanda. Using the method of constant stimuli, these were presented to subjects as a pair at tone durations ranging from 100 ms to 2.7 ms. The pairs, with one tone played 500 ms after the other had finished, came in the order 'high-low' or 'low-high'. They were backward- and forward-masked with white noise. Subjects were required to state, in their own time, the temporal order of the tones. The task was exploratory and imperfect—white noise was not effective as a mask and the 500 ms inter-tone gap allowed subjects to rehearse the first tone—but auditory IT correlated at  $-0.66$  with Mill Hill IQ

scores and at  $-0.70$  with scores on Raven's Progressive Matrices ( $n = 13$ ). Excluding two mentally handicapped subjects, the correlation with Mill Hill IQ remained significant at  $-0.61$ , but the correlation with Raven scores fell to a non-significant  $-0.28$ .

The correlation between visual and auditory IT was  $0.99$ , but it was entirely dependent upon the inclusion of the two mentally handicapped subjects. This study provided tentative evidence that IT might be a general, not solely visual, information processing or encoding speed. However, a series of objections to any such conclusion came in the study of Irwin (1984). Despite replicating the auditory IT-IQ correlation in a sample of 50 schoolchildren, Irwin noted the following:

- 1) Using a task very like that of Deary (1980), his subjects often obtained ITs at brief durations where there was overlap of frequency spectra for the stimulus tones, making the task, in effect, partly a pitch discrimination task.
- 2) There was a significant correlation between auditory IT and pitch discrimination ( $r = -0.54$ ).
- 3) There was a near-zero correlation between auditory and visual IT.

Clearly, Irwin's (1984) results argue against a general information-intake ability and against an auditory processing speed-IQ relationship.

There were problems with Irwin's study. Chiefly, he did not report having pretested, as Deary (1980) did, his subjects for pitch discrimination ability before including them in the auditory IT analysis. This is equivalent to not excluding subjects with poor visual acuity from a visual IT study. About one-third of normal adults and up to one-half of samples of 12-year-old children (Deary *et al.*, 1989) cannot reliably discriminate 880 from 770 Hz, however long the presentation time. If these subjects were not excluded from the auditory IT analyses there would be a spurious inflation of any existing pitch discrimination-auditory IT correlation.

Nettelbeck, Edwards and Vreugdenhil (1986) improved the auditory IT task by replacing the white noise mask with a mask consisting of rapidly alternating 15 ms bursts of both stimulus tones and white noise. The inter-tone gap remained and, perhaps owing to the non-exclusion of poor pitch discriminators, the authors obtained a very skewed (towards long ITs) distribution of auditory ITs. Nevertheless, in their sample ( $n = 29$ ) auditory and visual IT correlated significantly with Raven's Advanced Progressive Matrices (at  $-0.38$  and  $-0.41$ , respectively) and with each other ( $-0.39$ ).

Three recent experiments in our laboratory have addressed the above issues using a further improvement of the auditory IT task. To achieve mean IT values in a range that does not involve the overlap of the stimulus tones' frequency spectra, we have made two changes:

- a) the inter-tone gap has been removed; and
- b) we have introduced a more effective mask, consisting of alternating 10 ms bursts of the stimulus tones.

In contrast with previous studies, we obtain, when those subjects with poor

pitch discrimination skills are excluded, relatively unskewed distributions for auditory IT and mean values of about 70 ms for 90% accuracy levels.

In a study on undergraduates we found that auditory IT correlated significantly with Alice Heim 5 ( $n=40$ ,  $r=-0.31$ ) and Mill Hill Vocabulary scores ( $n=80$ ,  $r=-0.27$ ) but not with Raven's Advanced Progressive Matrices scores ( $n=80$ ,  $r=-0.05$ ) (Deary *et al.*, 1989). In the same study auditory IT estimates were correlated with three different IT tasks, some more vulnerable to strategy use than others, resulting in coefficients of 0.20 (ns), 0.24 ( $p < 0.05$ ) and 0.53 ( $p < 0.001$ ). This result, together with that of Nettelbeck, Edwards and Vreugdenhil (1986), supports the hypothesis that auditory and visual IT share common variance, at least in adults.

Using the same auditory IT task with a new sample of 34 undergraduates, we have replicated the correlation between auditory IT and Alice Heim 6 Verbal scores ( $-0.45$ ), but we found a non-significant correlation with Alice Heim 6 Non-verbal scores ( $-0.27$ ) (Deary, Head and Egan, 1989). In this group there was a near-zero correlation between Seashore Pitch Perception scores and cognitive ability, and a non-significant correlation between auditory IT and pitch perception scores ( $-0.20$ ). In a group of schoolchildren ( $n=53$ ) we found that the auditory IT correlated significantly with both Raven's Matrices ( $-0.26$ ) and Mill Hill Vocabulary scores ( $-0.36$ ). However, in the schoolchildren ( $n=119$ ) pitch discrimination ability did correlate significantly, but at levels around 0.17, with IQ-type test scores (Deary, Head and Egan, 1989). When partial correlations were calculated for auditory IT and IQ, controlling for pitch discrimination ability, they deviated very little from the original estimates.

There are still too few studies on auditory IT to draw firm conclusions, but the above evidence allows us to state the following interim hypotheses:

- 1) Auditory IT ability correlates significantly with psychometric intelligence test scores, particularly verbal test scores.
- 2) This correlation is not dependent upon pitch discrimination ability.
- 3) There is a moderate correlation between auditory and visual IT.

The correlations mentioned above are uncorrected for restriction of ability range in samples and for unreliability of the IT and IQ measures and are, therefore, underestimates of the true IT-IQ correlation. At this stage, then, it is still possible credibly to maintain that information encoding speed accounts for a moderate amount of IQ variance and that this encoding speed is not a property of just one sensory modality.

### EVOKED POTENTIAL STUDIES OF INSPECTION TIME

Posner (1986) makes the case that, in what he terms 'mental chronometry', analyses of averaged evoked potentials (AEPs) can provide converging evidence which is a useful adjunct to the methods of experimental psychology. We now discuss experiments which establish the existence of individual differences in AEP measures which can be related to the individual's IT, and to psychometric intelligence.

Earlier analysis of AEP-intelligence relationships involved relatively gross

measures (for example, Blinkhorn and Hendrickson's (1982) string-length measure) of response to simple, repetitive stimuli, where the task (e.g. simply listening) imposed minimal demands on the subjects. It is important to stress that our experiments are distinct from this earlier literature. Our subjects were faced with the equivalent of an IT-task, with presentation durations chosen to ensure 10% or 15% errors, and we are concerned with individual differences in latency, amplitude or speed of development of specific AEP peaks.

All experiments in this section involve basically simple visual discriminations, such as whether the longer of two lines (presented on a 7-segment LED display, with one line twice the length of the other) is on the left or the right. The task is made difficult by imposing, after a suitable interval, a backward mask.

It is conventional in analyses of evoked potentials to present the stimuli for a fixed duration. Because of individual differences in speed of discrimination, adoption of a constant stimulus duration before mask onset would produce a task which differed in difficulty for different subjects. To ensure that the features identified are not merely correlates of differences in task difficulty, we presented subjects in each experiment with discriminations of constant *psychological* difficulty, presenting stimuli for a duration equal to their (previously determined) inspection time.

Having chosen to match across subjects for psychological difficulty, we need to establish that the differences we report are not a trivial side effect of differences in physical duration. The first two experiments introduce the AEP measures of interest, and establish that they are unaffected by gross variation in physical duration of the stimuli used in the task.

Both experiments were small-sample studies (Zhang, Caryl and Deary, 1989b), in which subjects' IT (to a 90% correct criterion) had been tested immediately before the AEP session, using the stimuli to be presented in the AEP work. All testing took place in a darkened room, and stimuli were presented with a 7-segment LED display. Subjects were required to respond at leisure (and never before the mask was switched off) rather than to treat the experiment as a reaction time task. Speed of responding was unimportant, and was not recorded. Each trial began from two to three seconds after the previous response.

Silver-silver chloride electrodes were used. The active electrode at the vertex was referred to the left mastoid, with the right mastoid as earth. 1024 points following cue onset were sampled at 1 kHz.

In both experiments, response to stimuli at the subject's IT duration was contrasted with that to stimuli representing much easier or much harder discrimination tasks. In experiment 1, easy stimuli were presented for 1.75 times the subject's IT duration, before mask onset, difficult stimuli for only 0.25 times his IT. In experiment 2, easy stimuli were completely unmasked, while in the 'difficult' trials, the mask was presented synchronously with stimulus onset, so that the stimulus was completely obscured. In each experiment, subjects were presented with 75 trials at IT-duration, randomly intermixed with 70 easy and 70 difficult trials, and separate AEPs were generated for each of the three conditions, using the first 64 stimuli without visible artefacts. Measures used were P200 and P300 latencies and amplitudes, and a measure of the rise-time of the P200 wave (P200T) (Zhang, Caryl and Deary, 1989a).

The IT task requires a subject to encode rapidly the discriminative stimulus into STM, before onset of the mask; once encoding has taken place, the decision about which alternative was presented can be made at leisure. Chapman, McCary and Chapman (1978) identified a factorial component, peaking slightly later than the conventionally recognized P200 wave, whose size was correlated with degree of encoding of a discriminative stimulus into STM. This evidence, and pilot work from our laboratory, identified the P200 wave as one of particular interest. The P300 to stimuli of psychological significance is often considered to mark completion of decision-making, and provides an index of this which is unaffected by variation in time required for response selection (Pritchard, 1981). P300 amplitude should reflect confidence in the decision.

As expected, P300 amplitude did vary significantly between the three categories of trials in each experiment (repeated measures ANOVA,  $F(2,14) = 5.49$  in experiment 1 and  $4.34$  in experiment 2, both  $p < 0.05$ ). Page's test for trend confirmed that, in each experiment, P300 amplitude was greatest for the easy discrimination, intermediate for the IT-duration condition, and smallest for the condition where the discrimination was very difficult or impossible. In neither experiment was there significant variation in P300 latency across the three categories of trial ( $F(2,14) = 0.19$  and  $1.82$ , respectively). The contrast with experiments in which decision latency increases with difficulty (see, for example, McCarthy and Donchin, 1981; Duncan Johnson and Kopell, 1981) presumably reflects the fact that stimulus information remained available in these alternative paradigms, allowing further sampling where the discrimination was difficult, while in our experiments early masking eliminates this possibility.

Turning to the P200 measures, we found that the rise-time measure P200T was correlated with IT more highly and more consistently than other measures. Focusing initially on the IT-duration trials, which are of a constant difficulty for all subjects in both experiments, and combining data from the two experiments to avoid some of the problems of interpretation of results based on very small numbers, we found a correlation between P200T and IT of  $0.57$  ( $n = 16$ ,  $p < 0.05$ ).

This correlation is based on data in which higher-IT subjects receive longer stimulus presentations before mask onset. To find whether this variation in pre-mask stimulus duration might be important, we need to look for an effect of the seven-fold variation in stimulus duration across the three types of trial in experiment 1. A repeated-measures ANOVA test revealed no significant effect of presentation duration on any P200 measure ( $F(2,14) = 0.61$ ,  $0.71$  and  $1.96$  for P200T, P200A and P200L, respectively), but confirmed the presence of significant differences between subjects for all three measures ( $F(7,14) = 20.23$  for P200T and  $14.47$  for P200A, both  $p < 0.01$ , and  $3.27$ ,  $p < 0.05$ , for P200L). Thus individual differences, and not the variation in stimulus duration, appear to underlie the IT-P200T correlation.

Could we influence these P200 measures consistently by the more drastic manipulation of eliciting stimuli in experiment 2? Again, there were no significant differences between experimental conditions for any P200 measure ( $F(2,14) = 0.14$ ,  $0.67$  and  $2.12$  for P200T, P200A and P200L), but highly significant differences between subjects ( $F(7,14) = 6.84$ ,  $20.45$  and  $34.34$  for P200T, P200A



and P200L,  $p < 0.01$  or better in each case). Note that in both experiments, the AEP measure (P300A) which reflected ease of the discrimination did vary between conditions. We can safely conclude that the individual differences in P200T which we are interested in are not an artefact of differences in stimulus duration or the ease of the discrimination.

As might be expected, in view of the insensitivity of the P200T measure to major variation in presentation duration, we have also found that individual differences in this measure can be identified in responses to the variable-duration stimuli presented while the subject's IT is being measured using the PEST procedure, as well as with standard-length stimuli, and that here too the P200T measure is correlated with IT (Zhang, Caryl and Deary, 1989a).

If the correlation between IT and IQ is approximately 0.5 in the overall population, there must be a considerable task-specific component of the IT variance. We now argue as follows: Since IT and IQ correlate, and since we have demonstrated a correlation between P200T and IT, then:

- 1) Might the P200T index be a correlate of IQ?
- 2) Alternatively, might the P200T index be a non IQ-related (i.e. task-specific) component of IT variance?

We also wanted to know whether the difference in P200T we have described is present in potentials evoked only by the IT stimuli (requiring rapid encoding or discrimination), by any stimuli requiring discrimination (even if presented for several hundred milliseconds), or by any stimuli at all, no matter whether or not they require a response.

Experiment 3 involved a larger number of subjects, and addressed these questions (Zhang, Caryl and Deary, 1989b). The experimental procedure was more complex than in the previous experiments, and we shall summarize only the most relevant results.

Subjects had their IT measured in a pre-test session, and in the main session were presented with post-masked IT-duration stimuli, as before. Their response had to be delayed until a response signal was presented (on the same LED panel). In only half the trials was the subject required to attend to and discriminate the IT-duration stimuli; in the other half, the IT stimulus could be neglected, but to maintain attention to the LED panel, on these trials we required subjects to look for the response signal, and respond as rapidly as possible when it appeared, pressing a standard key for these RT responses. (Note that in this experiment rapid responding was also required on the trials in which IT-stimuli were *discriminated*, but that two response keys were used for this discriminative response). The pre-stimulus cue was used both to alert the subject for the trial, and to signal whether the trial to follow required an IT-discrimination, or an RT response to the response signal. The pre-stimulus cues used were the digits 2 and 6, also presented on the LED panel. Evoked responses to the cues, and to the IT stimuli (whether or not these were to be discriminated) were recorded. Evoked responses to the response cues were not recorded. The meaning of the different cue digits, order of presentation of trials, etc., were appropriately randomized.

We shall deal here with results for the vertex electrode only (Table 16.1). In



Table 16.1 Correlations of AEP measures, IT and IQ

AEP measure	Stimulus	Correlation with	
		IT ( <i>n</i> = 37)	IQ ( <i>n</i> = 35)
P200 <sub>T</sub>	IT(discriminated)	0.645 <sup>c</sup>	0.111
	IT(ignored)	0.117	0.047
	Cue	0.290 <sup>d</sup>	-0.339 <sup>a</sup>
P200 <sub>L</sub>	IT(discriminated)	0.442 <sup>b</sup>	-0.017
	IT(ignored)	0.070	0.160
	Cue	-0.003	0.119
P200 <sub>A</sub>	IT(discriminated)	0.332 <sup>a</sup>	0.101
	IT(ignored)	0.253	-0.095
	Cue	-0.000	0.195
P300 <sub>A</sub>	IT(discriminated)	0.503 <sup>b</sup>	0.173
	IT(ignored)	0.068	0.069
P300 <sub>L</sub>	IT(discriminated)	0.055	0.066
	IT(ignored)	0.061	-0.062

IT(discriminated) represents the responses to IT stimuli to which the subject was required to make a discriminative response. IT(ignored) represents the responses to IT stimuli which the subject could ignore, being required to make a RT response but no discrimination. Cue represents responses to the digit used both as a cue and to indicate whether a RT or discriminative response to the subsequent IT stimulus was required.

<sup>a</sup> *p* = 0.05; <sup>b</sup> *p* = 0.01; <sup>c</sup> *p* = 0.001; <sup>d</sup> *p* = 0.05 (one-tailed).

this experiment there were significant correlations between IT and the P200T of potentials elicited both by IT stimuli on trials requiring discrimination, and by the cues. In contrast, potentials evoked by the IT stimulus on trials in which it could be disregarded had a P200T which was not correlated with IT. This correlation is evidently dependent on the need to encode and subsequently discriminate the stimulus (as was true both for the cues, and for task-related IT stimuli), but does not depend on the speed of encoding required with the IT stimuli.

In contrast to the previous experiment, there were also correlations between the P200L and P200A measures and IT. These were only obtained for responses to the IT-stimuli which were to be discriminated; for these stimuli, P300A was also significantly correlated with IT.

Do these relationships reveal anything about psychometric intelligence? The P200T measure of responses evoked by the cue (but not the IT stimulus) was correlated with AH5 total score ( $r = 0.34$ ,  $p < 0.05$ ). The sign of this correlation is as expected in view of the IT-IQ relationship. The correlation was stronger with part I of the AH5 test (verbal-mathematical) rather than part II (pictorial-spatial). None of the other evoked potential measures which correlated with IT correlated with intelligence.

The pattern of relationships can be most conveniently summarized by a small principal components analysis of the correlations (Table 16.2), which reveals three factors, each extracting approximately 24% of the variance, the third of which (with loadings on IT, AH5, and the P200T to the cue) presumably reflects

Table 16.2 Principal components analysis of selected AEP measures, IT and IQ (based on correlations for 35 subjects)

Measure	Sorted factor loadings		
	Factor 1	Factor 2	Factor 3
P200 <sub>L</sub>	0.923	-0.102	0.068
P200 <sub>T</sub>	0.872	0.300	0.013
IT	0.611	0.496	-0.469
P300 <sub>A</sub>	0.151	0.876	0.066
P200 <sub>A</sub>	0.034	0.874	0.031
AH5 (part 1)	0.117	0.197	0.846
P200 <sub>T</sub> (to cue)	0.119	0.087	-0.687
AH5 (part II)	0.021	0.008	0.644
Variance	25%	24%	23%

Except where specified, all AEP measures are for responses evoked by IT stimuli which were to be discriminated by the subject.

general intelligence. The first and second factors, which have loadings on IT and evoked potential measures but not on intelligence, must reflect task-specific components of the variance.

In summary, our results suggest that differences in IT performance depend on individual differences in stimulus processing, which can be detected as early as 200 ms after stimulus onset, in responses to unmasked as well as the conventional backward-masked stimuli used in IT tests, provided the stimulus must be encoded. The differences are apparently not shown to stimuli which can be neglected. Our AEP measures revealed an important task-specific component in responses to IT stimuli, as well as a non-specific component related to general intelligence.

## REFERENCES

- Beck, L. F. (1933). The role of speed in intelligence. *Psychological Bulletin*, **30**, 169-178.
- Blinkhorn, S.F., & Hendrickson, D. E. (1982). Average evoked responses and psychometric intelligence. *Nature*, **295**, 596-597.
- Brand, C. R., & Deary, I. J. (1982). Intelligence and 'inspection time'. In H.J. Eysenck (ed.), *A model for intelligence*. New York: Springer.
- Burt, C. R. (1909/10). Experimental tests of general intelligence. *British Journal of Psychology*, **3**, 94-117.
- Cattell, J. McK. (1886). The inertia of eye and brain. *Brain*, **8**, 295-381.
- Cattell, J. McK. (1890). Mental tests and measurements. *Mind*, **15**, 373-381.
- Chapman, R. M., McCary, J. W., & Chapman, J. A. (1978). Short-term memory: the 'storage' component of the human response predicts recall. *Science*, **202**, 1211-1214.
- Deary, I. J. (1980). How general is the mental speed factor in 'general intelligence'? B.Sc. Hons. thesis, University of Edinburgh.
- Deary, I. J. (1986). Inspection time: discovery or rediscovery? *Personality and Individual Differences*, **7**, 625-631.

- Deary, I. J. (1988). Basic processes in human intelligence. In H. J. Jerison and I. Jerison (eds.), *Intelligence and evolutionary biology*. Berlin: Springer.
- Deary, I. J., Caryl, P. G., Egan, V., & Wight, D. (1989). Visual and auditory inspection time: their interrelationship and correlations with IQ in high ability subjects. *Personality and Individual Differences*, **10**, 525-534.
- Deary, I. J., Head, B., & Egan, V. (1989). Auditory inspection time, intelligence and pitch discrimination. *Intelligence*, **13**, 135-148.
- Duncan Johnson, C. C., & Kopell, B. S. (1981). The Stroop effect: brain potentials localise the source of interference. *Science*, **214**, 938-941.
- Hunt, E. (1980). Intelligence as an information processing concept. *British Journal of Psychology*, **71**, 449-474.
- Irwin, R. J. (1984). Inspection time and its relation to intelligence. *Intelligence*, **8**, 47-65.
- Jensen, A. R. (1985). The plasticity of 'intelligence' at different levels of analysis. In J. Bishop & D. Perkins, (eds.), *Thinking: Progress in research and teaching*. Franklin Institute Press.
- Jensen, A. R., & Vernon, P. A. (1986). Jensen's reaction time studies: a reply to Longstreth, *Intelligence*, **10**, 153-179.
- McCarthy, G., & Donchin, E. (1981). A metric for thought: a comparison of P300 latency and reaction time. *Science*, **211**, 77-80.
- Mackenzie, B., & Bingham, E. (1985). IQ, inspection time and response strategies in a university population. *Australian Journal of Psychology*, **37**, 257-268.
- Mackenzie, B., & Cumming, S. (1986). Inspection time and apparent motion. *Personality and Individual Differences*, **7**, 721-729.
- Mackintosh, N. J. (1981). A new measure of intelligence? *Nature*, **289**, 529-530.
- Nettelbeck, T. (1982). Inspection time: an index for intelligence? *Quarterly Journal of Experimental Psychology*, **34A**, 299-312.
- Nettelbeck, T. (1987). Inspection time and intelligence. In P. A. Vernon (ed.), *Speed of information processing and intelligence*. Norwood, NJ: Ablex.
- Nettelbeck, T., Edwards, C., & Vreugdenhil, A. (1986). Inspection time and IQ: evidence for a mental speed-ability association. *Personality and Individual Differences*, **7**, 633-641.
- Nettelbeck, T., & Lally, M. (1976). Inspection time and measured intelligence. *British Journal of Psychology*, **67**, 17-22.
- Posner, M. I. (1986). *Chronometric explorations of mind*. Oxford: Oxford University Press.
- Pritchard, W. S. (1981). Psychophysiology of P300. *Psychological Bulletin*, **89**, 506-540.
- Sternberg, R. J., Conway, B. E., Ketron, B. L., & Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology*, **41**, 37-55.
- Vickers, D., Nettelbeck, T., & Willson, R. J. (1972). Perceptual indices of performance: the measurement of 'inspection time' and 'noise' in the visual system. *Perception*, **1**, 263-295.
- Wissler, C. (1901). The correlation of mental and physical tests. *Psychological Review Monographs*, **3**.
- Zhang, Y., Caryl, P. G., & Deary, I. J. (1989a). Evoked potential correlates of inspection time. *Personality and Individual Differences*, **10**, 379-384.
- Zhang, Y., Caryl, P. G., & Deary, I. J. (1989b). Evoked potentials, inspection time and intelligence. *Personality and Individual Differences*, **10**, 1079-1094.